

State *or* Nature? Formal vs. Informal Sanctioning in the Voluntary Provision of Public Goods^{*}

Kenju Kamei^a, Louis Putterman^a and Jean-Robert Tyran^b

^a Department of Economics, Brown University, Providence, RI 02912, United States

^b Department of Economics, University of Vienna, Vienna, Austria and Department of Economics, University
of Copenhagen, Denmark.

Abstract

The sanctioning of norm-violating behavior by an effective formal authority is an efficient solution for social dilemmas. It is in the self-interest of voters and is often favorably contrasted with letting citizens take punishment into their own hands. Allowing informal sanctions, by contrast, not only comes with a danger that punishments will be misapplied, but also should have no efficiency benefit under standard assumptions of self-interested agents. We experimentally investigate the relative effectiveness of formal vs. informal sanctions in the voluntary provision of public goods. Unsurprisingly, we find that effective formal sanctions are popular and efficient when they are free to impose. Surprisingly, we find that informal sanctions are often more popular and more efficient when effective formal sanctions entail a modest cost. The reason is that informal sanctions achieve more efficient outcomes than theory predicts, especially when the mechanism is chosen by voting.

JEL classification codes: C92, C91, D03, D71, H41.

Keywords: Sanction, social dilemma, public goods, voluntary contribution mechanism, punishment, experiment.

^{*} We wish to thank the Danish research council (FSE) for financial support under project title “Cooperation and Institutions.” We thank participants at the Economic Science Association meeting in July, 2010 in Copenhagen for helpful comments.

1. Introduction

Although economics devotes much of its attention to situations in which the pursuit of self-interest by each leads to greater social welfare for all, social dilemmas, or situations in which universally self-interested behavior makes a social good less rather than more attainable, are an almost equally common occurrence in social life. When individuals follow the logic of rational self-interest and refuse to voluntarily contribute to the provision of public goods, recourse is often had to the coercive powers of the state, which can make contributions (such as tax payments) mandatory and subject to material and other penalties for non-compliance. But there are numerous collective action dilemmas for which this solution seems unavailable, including public goods that fall beneath government notice because they are too local or are valued by only a specific sub-set of citizens. Moreover, the very existence of a non-authoritarian government that can promulgate and enforce regulations to deter free riding depends on such voluntary pro-social acts as citizen scrutiny of politicians' actions, self-education about political issues, and making the effort to vote in elections. The relationship between voluntary and coerced collective action is thus a complex one, with the question of when we can rely upon formal sanctions to resolve social dilemmas and when voluntary cooperation is desired an important and underexplored issue.

Economists studying public goods problems from theoretical and practical standpoints have been joined in recent decades by experimentalists studying aspects of the topic in controlled laboratory settings. One particularly active area of research of late has been studies of the possibility of addressing voluntary collective action problems by allowing individual members of a group to impose monetary sanctions on other members at their own discretion and expense (Ostrom, Walker and Gardner, 1992; Fehr and Gächter, 2000; Masclet *et al.*, 2003; Page, Putterman and Unel, 2005; Sefton, Shupp and Walker, 2007).¹ This work is of considerable interest both theoretically, since it suggests inclinations to incur costs to punish that are absent in standard models, and practically, for instance with respect to management of commonly held resources and addressing problems of effort in partnerships and teams. But it is difficult to think of real-world decentralized

¹ For additional references and a comprehensive overview, see Herrmann and Gächter, 2008, and Chaudhuri, forthcoming.

monetary sanctioning regimes, so their use in experiments may be best thought of as stand-ins for social penalties (condemnation, ostracism, loss of esteem). A more important point to note is the curious fact that comparable attention has not been paid to more centralized forms of governance and to the conditions, if any, under which informal sanctions regimes are preferable to formal ones.

In a formal sanctions regime, a group adopts a rule specifying what penalties will be imposed under what conditions and sets up a body (in large group settings, an administration or government) that observes rule violations and imposes the stipulated penalties. Whereas it is never privately optimal for a rational decision-maker concerned only with his/her own material payoff to engage in costly sanctioning in a world of common knowledge that all are of that type, it is quite rational for such an individual to vote for a system of centrally-imposed sanctions capable of making cooperation the privately optimal choice of all, as long as the cost doesn't exceed the benefit. Even when rationality, self-interest, and common knowledge assumptions are dropped, a formal sanctioning regime may be preferred because there may be near certainty that a formal sanction will be implemented but uncertainty about who will impose informal sanctions and when. The predictability of formal sanctions may mean that fewer costly sanctions need be imposed under a formal sanctions regime, an efficiency advantage. If informal sanctions do occur when allowed, there may be no way to guarantee that they are directed at the right individuals, whereas rules in the interest of every group member can assure that formal sanctions are well targeted. This may explain why the centralized administration of penalties has been seen for at least three thousand years as the hallmark of civilization, whereas the enforcement of rules by individuals is often denigrated as "vigilante justice" or "mob rule."

There are also reasons for preferring informal sanctions in some circumstances, however. We think of informal sanctions in connection with village management of woodlots and irrigation systems or social pressure in work teams, while formal sanctions bring to mind managerial structures and governments. Having an enforcing body in place—for example, a police force and courts—involves fixed costs that, paradoxically, may need to be used relatively little if their presence suffices to deter the social bads they are meant to prevent; yet those fixed costs may be large. Informal sanctions may be less deterrent

because they are less certain, but they may have the advantage of avoiding fixed administrative costs.²

We conducted experiments allowing us to compare the operation of formal and informal sanction schemes in the n-person collective action problem known as the voluntary contribution mechanism (VCM) or public goods game (PGG). Along with a companion paper (Markussen, Putterman and Tyran, 2011), we believe ours to be the first experimental study in which subjects choose between formal and informal sanctions. *Inter alia*, our experiment differs from Markussen *et al.* in that if our subjects choose to use a formal scheme, they also must select its parameters. (In Markussen *et al.*, subjects vote for or against pre-packaged schemes made available by the experimenter.) We use three related experimental designs to study choice of and behavior under formal versus informal schemes: two designs with and one without prior exogenous experience of sanction schemes, where the two designs with prior experience differ in the order in which formal and informal sanctions schemes are introduced. Each of the three designs is implemented both in sessions in which formal sanction schemes entail a fixed administrative cost, and in sessions in which they do not, yielding a total of six treatments. We also conduct sessions with the same underlying voluntary contribution structure but in which neither formal nor informal sanctions are available, so as to provide a baseline against which to measure the effects of sanctions on cooperation and efficiency.

Our experiments largely confirm standard theory's prediction that subjects who control details of a formal sanctions scheme will choose an efficient one, sanctioning free riding on contributions and doing so at rates rendering socially optimal contributing privately optimal. However, while standard theory predicts a preference for formal over informal sanctions whenever the former cost less than the gains associated with the shift from full free riding to socially optimal contributing, most of our subjects in fact vote for informal rather than formal sanctions when formal sanctions cost as little as a quarter of

² In some circumstances, a government may promulgate a formal prohibition backed by penalties but may lack the enforcement resources to catch and penalize more than a handful of violators, whereas if most members of the local community embrace the goal in question, informal sanctioning may be a high probability response to norm violation. In such a case, the assumption that there is greater certainty regarding a formal than an informal sanction should be reversed.

those gains. This is explained, in part, by the fact that our subjects, like those in other experiments, engage in costly punishment directed mainly at low contributors when they operate under an informal sanctions scheme, thus inducing increased contributions to the public good in the absence of formal sanctions and without incurring a fixed administrative cost. In this sense, our results reinforce the growing body of evidence for the existence of propensities that render self-governance more achievable than standard economic theory implies. But informal sanctions are used even more efficiently by our subjects than by those in many other experiments, with some indication that the endogeneity of the sanctions regime plays a part.

Looking beyond our experimental set-up, for societies taken as wholes, a state or central administrative body having the power to enforce cooperation and the “bottom up” voluntary cooperation that is facilitated by informal sanctions (the “Nature” we counter-pose to “State” in our title) might best be thought of less as alternatives to each other than as complements. A well-functioning state that acts dispassionately in the general interest may be impossible unless undergirded by some amount of voluntary collective action, e.g. willingness to forego short-term self-interest on the parts of at least some public servants, and of an involved and vigilant populace.³ Although we don’t explicitly make voluntary collective action an input to or a prerequisite for the operation of formal sanctions in the experimental designs studied here, the capacities that our subjects display for cooperating under informal sanction regimes can be viewed as good rather than bad news for formal institutions, since they suggest that the voluntarily cooperative substrate required by good formal institutions is attainable given the richness of actual human behavior, whereas it appears to be out of reach in a theoretical world of *Homo economicus*.

The rest of the paper proceeds as follows. In Section 2, we discuss theoretical considerations regarding collective action and formal and informal sanctions, and we briefly review relevant experimental research. Section 3 spells out our experimental design, the

³ As James Madison wrote “If there be no virtue among us, no form of government can render us secure. To suppose that any form of government will secure liberty or happiness without any virtue in the people is an illusion.” (Speech to Virginia Ratifying Convention, June 20, 1788). The same theme appears almost two millennia earlier in the observation of Horace that “Leges sine moribus vanae” (Laws without morality are useless). The latter is cited by Gintis (2000, p. 116).

rationale behind the various treatments, and theoretical predictions both under the assumption of common knowledge of rationality and self-interest and under alternative behavioral assumptions. Section 4 summarizes the experimental results and provides analysis of key relations within our data. Section 5 concludes the paper with a brief summary and suggestions for future research.

2. Theory and literature

We consider a group with n members who engage in a finitely repeated interaction where, in each period, each individual must allocate an endowment of E_i money units between a private and a group account. Money placed in the group account is scaled up and divided equally among the members, generating a marginal private return on contributions m (where $0 < m < 1$) and a return to the group of mn (where $1 < mn < n$). In this canonical voluntary contribution mechanism (VCM) or public goods game (PGG), total earnings are maximized if each member places the full E_i in the group account, but it is in each individual's private interest to put zero in that account.⁴ Although cooperation might be established among rational selfish individuals, were the stage game to be infinitely repeated, presence of a known last period eliminates this possibility due to the usual backward induction argument. If we depart from the assumption that all members are strictly self-interested, rational, and have common knowledge of this, cooperation may be possible (Kreps *et al.*, 1982); we discuss such departures later.

a) Behavior under and voting on sanctions regimes, according to standard theory

Suppose that group members are asked to vote for one of two alternatives: (a) engaging in the interaction under the rules described above, and (b) implementing a formal sanction scheme where for each money unit an individual puts in their private account, she is fined s units, with $s > 1 - m$, and implementing the scheme costs each individual f per period. Then as long as $f < (mn - 1) \cdot E_i$ and the assumption of common knowledge of self-interest and rationality holds, group members can expect to earn more per period with than without the

⁴ A version of the VCM with interior social optimum entailing contributions to both private and group accounts is more pertinent to many policy settings, but the difference is inessential to our theoretical interests and a set-up with interior equilibrium introduces computational complications in laboratory implementation that we chose to avoid.

scheme. The reason is that the scheme induces a rational individual to contribute E_i to the group account, and with all doing this, each earns $mnE_i - f$ per period, which exceeds the E_i per period earned under option (a). An individual would accordingly choose (b) over (a) if expecting to influence the outcome. While a single individual who has reason to believe that her vote cannot be pivotal will be indifferent regarding which scheme to vote for, we assume that coordination, trembling-hand perfection, or similar arguments lead to choice of the weakly dominant alternative (b) by the group.⁵

Suppose now that there is a third option, (c), wherein the game played by the group of individuals each period has two stages, a first stage in which each individual decides how many of his or her E_i units to allocate to the group account, and a second stage in which each learns the amount allocated by the other $n - 1$ individuals and can reduce the earnings of any other group member by P units per unit deducted from her own provisional earnings for the period. An individual's earnings for a period are therefore given by

$$(E_i - C_i) + m \cdot \sum_{j \text{ includes } i} C_j - \sum_{\text{all } j \neq i} p_{ij} - P \cdot \sum_{\text{all } j \neq i} p_{ji} \quad (1)$$

where p_{ij} indicates the number of units of punishment that group member i gives to member j and conversely for p_{ji} , so that the third term is i 's expenditure on punishing others and the fourth is i 's losses from being punished by others. In this case, a credible threat of receiving $P > (1 - m)$ units of punishment for each unit of endowment placed in one's private account can make it privately optimal to contribute to the group account. However, in a finitely repeated game with rational, strictly self-interested agents having common knowledge that all are of the same type, there can be no such credible threat, since it is never rational to punish in the last period and therefore punishing in any earlier period as a way of threatening to punish future free riding is not credible. For this reason, any of a set of members with common knowledge of rationality and payoff maximizing preferences can expect the availability of a punishment option to make no difference, the outcome still being uniform free riding and universal earnings of E_i per period. Assuming uniform voting for

⁵ In our experiment, there is no opportunity for subjects to communicate before voting. Therefore, a rational subject can never confidently put at zero the chance that her vote will be pivotal, and since there is no cost of voting, differential or otherwise, it seems reasonable to assume that the rational voter will vote for her preferred option. On trembling hand perfection, see Selten, 1975.

the weakly dominant alternative, then, standard theory predicts that option (b) will be selected in favor of either of options (c) and (a), while individuals will be indifferent between options (a) and (c). (In our experiment, subjects will vote only between (b) and (c), so the operational prediction is that they will vote for (b).)

b) Possible behavioral departures

Observations from past public goods experiments suggest that standard theory does poorly in predicting many decisions in settings of types (a) (simple VCM) and (c) (VCM with punishment stage). In dozens of experiments, subjects contribute 40 to 60% of their endowments to the group account on average in the initial period of finitely repeated play in a group of fixed but anonymous membership (Zelmer, 2003). Contributions trend downwards with repetition, but are rarely uniformly zero in the known last period. If previously unexpected additional rounds of play are announced and commence, there is a resurgence of contributions (Andreoni, 1988). There is strong evidence of reciprocal altruism or a “conditional willingness to cooperate” on the parts of many subjects (Fischbacher and Gächter, 2010 and Thöni *et al.* 2009). In experiments of type (c), costly punishment is given at some point in a repeated play treatment by the majority of subjects, and the average subject engages in costly punishment on some one-fifth to one-third of opportunities do so (Falk, Fehr and Fischbacher, 2005 and sources cited above). While subjects mainly direct punishment at the lower contributors in their groups, many subject pools also exhibit substantial punishment of high contributors (Cinyabuguma *et al.*, 2006; Herrmann *et al.*, 2008), which seems in part to be explained by “blind revenge.”⁶ Contrary to the prediction above that the availability of punishment does not alter equilibrium play, its availability at sufficiently low cost is associated with at least sustained if not rising contributions to the group account in numerous experiments (see, e.g., Nikiforakis and Normann, 2008; see also the discussion in Chaudhuri, forthcoming).

⁶ Low contributors who have received punishment seem to presume it has come from a higher contributor, with more than half of observed incidents of punishing high contributors coming in a period after the punisher has herself received punishment. The revenge is “blind” because the treatments in question leave the identity of the punisher anonymous. When subjects have opportunities to retaliate against a known punisher, there is considerable retaliation and the outcome is far less efficient (Nikiforakis, 2008; Denant-Boemont *et al.*, 2007), a problem to which we return toward the end of the paper.

While less studied, the observed impact of deterrent-strength formal sanctions on contributions appears largely consistent with the standard theoretical prediction (Putterman *et al.*, 2010). But two other aspects of formal sanctions schemes merit comment. First, subjects significantly increase their contributions in response to non-deterrent sanctions when these have been selected by the subjects themselves (Tyran and Feld, 2006), although equilibrium behavior under common knowledge and self-interest remains at zero contribution. Second, when choosing the details of a formal sanction scheme, some subjects fail to vote for imposing deterrent sanctions or for penalizing the socially inefficient action of contributing to private accounts, and such voting is correlated with self-assessed political conservatism and opposition to government intervention (Putterman *et al.*, 2010).

A final observation that may be pertinent to predicting behavior in our experiment is that the selection of an institution by voting may in and of itself influence its performance, as in the result of Tyran and Feld (2006). Dal Bó, Foster and Putterman (2010) demonstrate this using a novel design in a prisoners' dilemma experiment where the institutional change, that can be introduced either by vote or by an exogenous shock, is a unilateral punishment for defection, a change rendering mutual cooperation one of two equilibria (see also Sutter *et al.* 2010). Besides enhancing the likelihood that non-deterrent sanctions, if selected, would have an impact on contributions and earnings, this factor could also lead performance under informal sanctions to be more optimal under voted than under exogenous implementation. For example, a group vote to make informal sanctions available might be interpreted as a signal that most members are willing to punish free-riding and that they look favorably on others doing so, which might cause increased deterrence for given sanctions as well as better targeting of those sanctions that are imposed.

These observations affect our expectations about institutional choice in at least three ways. First, if efficiency is not as low in the absence of sanctions as is predicted by standard theory, then the cost that a rational subject should be willing to pay to have a formal sanction scheme in place is less than the amount predicted above $((mn - 1) \cdot E_i)$. Second, if availability of an informal sanction scheme generates a substantial increase in efficiency when the scheme is selected by vote, then subjects may prefer informal to formal sanctions when the latter entail a substantial fixed cost. Third, when groups choose to use formal

sanctions, we cannot take for granted that they will select the most efficient elements for their formal sanction scheme, and it will be interesting to look for variations in voting on the sanction parameters, variations that might be associated with individual propensities to contribute to the public good, with ideology, and perhaps with intelligence.

c) Literature

An early discussion of the state containing the germ of the idea that it solves a collective action problem is that of Thomas Hobbes (1996 [1651]). John Locke (2005 [1739]) explicitly considered society's choice between relying on decentralized (or horizontal) and using centralized (or vertical) punishment. Locke argued that assigning the rights of punishment to a central authority is necessary due to the self-oriented bias of individuals, and his arguments puts considerable weight on the danger of retaliatory punishment:

In the state of nature there often wants power to back and support the sentence when right, and to give it due execution. ... such resistance many times makes the punishment dangerous, and frequently destructive, to those who attempt it. (§. 126.)

... The inconveniencies that they are ... exposed to ... make them take sanctuary under the established laws of government ... [and] willingly give up every one his single power of punishing, to be exercised by such alone, as shall be appointed to it amongst them; and by such rules as the community ... shall agree on. (§. 127.)

In more recent political philosophy, the problem is addressed by Buchanan and Tullock (1962) who, in keeping with their skepticism regarding state intervention, take the more pragmatic view that voluntary cooperation is sometimes a superior option.

Experimental studies of endogenous choice of institutions for collective action began with work of Ostrom, Walker and Gardner (1992), who let subjects communicate and decide whether to use sanctions on overharvesting to achieve greater efficiency in a common pool resource game. More recently, Botelho *et al.* (2005), Gülerk *et al.* (2006), Ertan *et al.* (2009), and Sutter *et al.* (2010) have studied whether subjects will chose to face a voluntary contribution dilemma that permits or one that does not permit informal sanctions. Botelho *et al.* and Sutter *et al.* implement experimental designs in which subjects make a single choice over using or not using an informal sanction mechanism (and, in Sutter *et al.* also informal rewards). Both papers find informal sanctions to be unpopular. Gülerk

et al. and Ertan *et al.*'s designs include a series of decisions between allowing and not allowing informal sanctions, and both find that informal sanctions are unpopular at the outset but become increasingly popular with experience.⁷

Besides the present paper and Markussen *et al.*, we know of only two papers that study experimentally voting on formal sanctions in a VCM. Tyran and Feld (2006) let subjects vote on whether to impose on themselves a “mild law”—i.e., non-deterrent sanction—and compare a treatment in which the same sanction is imposed or not imposed exogenously. Some 60% of groups offered the choice choose the mild law and achieve higher efficiency than either those in the exogenous or endogenous no law and in the exogenous mild law conditions. Putterman *et al.* (2010) study a treatment to which an exogenous formal sanctions framework is assigned, but it is left to subjects to determine whether it is contributions to the group or to the private accounts that are subject to sanctions. Their subjects also vote on what the maximum sanction will be and what contribution level, if any, is exempt from sanctions. Seven of eight groups select fully efficient sanction schemes by their third of five votes.

In both the present paper and Markussen *et al.*, subjects have opportunities to vote on the use of sanction schemes, but the designs studied differ in several respects. First, in this paper, subjects' choice is always and only between formal and informal sanction schemes, whereas Markussen *et al.* also study choices between formal and no sanctions and between informal and no sanctions. Second, after our subjects decide to use a formal sanction scheme, they also decide by voting which allocation is sanctioned and at what rate, whereas in Markussen *et al.* each treatment comes with a “pre-packaged” formal sanction scheme, eliminating potential strategic voting issues. Third, a group adopting a formal sanctions scheme pays both a fixed cost and a cost per unit of sanctions imposed, in this paper, but a fixed cost only in Markussen *et al.* Finally, while parameters E (endowment), n

⁷ There are various differences between their designs, for example Botelho *et al.*'s subjects vote only before the experiment's final period after exogenously experiencing both no sanctions and informal sanctions regimes, while Sutter *et al.*'s subjects vote only before play begins. Gülerk *et al.*'s subjects “vote” by individually making a series of decisions about which of two groups, playing under different rules, to join. Ertan *et al.*'s subjects vote, like those in the other two papers, but over three ballot items that let them decide separately whether to allow punishment of below average, of above average, and of average contributors in their groups.

(group size), and m (private return from the public good) are the same in the two experiments, the possible fixed costs are 0 or 5 in this paper's treatments versus 2 and 8 in those of Markussen *et al.* Thus, subject responses to a larger range of costs (ranging from 0 to 40% of the earnings difference between full and no cooperation) can be studied by considering common elements of the two papers. We discuss how Markussen *et al.*'s findings complement ours in the paper's conclusion.

3. Experimental Design

Our experiment consists of six VCM (PGG) treatments in which subjects have opportunities to choose between informal and formal sanction schemes, with these voting opportunities either beginning at the outset or following exogenously assigned experience under each scheme. The experiment also includes a main control treatment without sanctions and a special control treatment for investigating whether the effect of informal sanctions is altered by endogeneity (see Table 1).

In all treatments, groups consist of $n = 5$ subjects (partner matching). Each subject is provided at the beginning of each period with a fixed endowment of $E = 20$ points (the experimental currency unit, redeemed at the ratio 34 points = 1 dollar at the session's conclusion), and is asked to make a series of decisions on the allocation of that endowment between a private and a group account. The marginal per capita return (MPCR) is $m = 0.4$, which makes contributing nothing to the group account the dominant strategy in each stage game, while contributing everything is the social optimum. We set the number of regular periods at 24 to allow for the evolution of group-level as well as an individual-level choice of institutions over time, and we group the periods into six phases of four periods each (with an additional, initial one period phase in the treatments in which voting on institutions begins without exogenous experience of sanction schemes). In the control or BASELINE treatment, subjects play the standard linear voluntary contribution game for all six phases, each separated by a break of 40 seconds to parallel the breaks for voting in the other treatments.⁸ The stage game is referred to as the No Sanctions (NS) regime, and under it

⁸ One of the effects of the voting breaks in the other treatments may be a "restart effect" (Andreoni, 1988), i.e. contributions tend to increase following any break in play. Inserting breaks into the BASELINE treatment thus puts it on a more level plain with the other treatments.

subjects learn at the end of each period the amount allocated to the group account by each of the others in their group in a random order. In the BASELINE treatment, all instructions occur before Phase 1, and subjects are aware that the same rules will apply for all 24 periods.

In the six voting treatments, subjects initially play for at least one period in the NS regime but subsequently play under either a Formal Sanctions (FS) or an Informal Sanctions (IS) regime, with the choice of regime either from the outset (6-vote treatments) or in later phases (3-Vote treatments) being made by majority vote. The upper four treatments displayed in Table 1 are ones in which subjects play their first three phases under exogenously imposed condition NS followed by either IS and FS (dubbed IF order, upper two cells) or FS and IS (FI order, next two cells). Subjects then play each of the last three phases under their group's majority choice—determined each time by a start-of-phase vote—of either the FS or the IS regime. In treatments in the left column, the FS condition carries no fixed administrative cost, whereas in those in the right column, each group member is charged 5 points per period—i.e., $f = 5$ —when the FS scheme is used. Distinctions by treatment order and administrative cost (none = N, 5 point cost = C) lead to our adopting the treatment names indicated in the four upper cells: 3(IF)-N, 3(IF)-C, 3(FI)-N and 3(FI)-C. In the two 6-Vote treatments, represented in the third row of cells, subjects play a single period in NS condition followed by six four-period phases, labeled Phases 2 – 7, played under each group's majority choice of either FS or IS, also determined in start-of-phase votes. Since there are no order differences among the 6-Vote treatments, they are distinguished only as 6-N and 6-C. The exogenous IS comparison treatment is discussed later.

In 3-Vote treatments, subjects learn about the conditions in four distinct sets of instructions. Before Phase 1, they learn about the VCM without sanctions; before Phase 2, they learn about either the IS or the FS condition, depending on order; and before Phase 3, they learn about the remaining condition. Before Phase 4, they read a brief set of instructions informing them about the voting that will take place at the beginning of each of the remaining three phases.⁹ In 6-Vote treatments, subjects learn about the VCM without

⁹ In all treatments, subjects are told at the outset that there will be six phases of four periods each and that they will always interact in the same group of five anonymous participants.

sanctions before their initial one-period phase (which for convenience of alignment with other treatments will be called Phase 0), then receive the remaining instructions about IS, FS, and votes between the two before their first four-period phase, here called Phase 1.¹⁰ In both 3-Vote and 6-Vote treatments, voting is simultaneous, mandatory, and free, with each subject indicating a preference for either IS or FS and groups learning their majority's preference but not the number voting for it. We conducted sessions with both 3-Vote and 6-Vote designs to check the robustness of voting outcomes and schemes' efficiencies to different opportunities for learning, and we conducted 3-Vote sessions with both IF and FI orders to test whether order matters when initial learning is possible. At least one period of NS play was included at the beginning in all designs so that subjects could learn about the basic voluntary contribution problem before the introduction of additional elements.

When IS is in place, each period includes both the initial contribution stage and a second stage in which subjects select a number of points of punishment, the minimum possible being 0 and the maximum 10, to assign to each other group member, the latter's contribution decisions being indicated in a random order. Each point of punishment assigned costs the recipient four points ($P = 4$), except if received punishment exceeds first stage earnings, in which case first stage earnings minus received punishment is set to 0. Period earnings could nevertheless become negative because the cost of imposing punishment is always incurred, with any one period losses being taken from the accumulated earnings of other periods.¹¹ Thus, period earnings under IS are given by

¹⁰ The instructions read by subjects referred to the phases by the numbers 1 – 7 rather than 0 – 6.

¹¹ The maximum of 10 points of punishment is common to many VCM-with-punishment experiments beginning with Fehr and Gächter (2000). In practice, of the 1090 instances in which a subject i punished another subject j , only 5 (0.46%) involved giving 10 points of punishment. We limited the effect of punishment so that period earnings could not become too negative, a matter of concern due to our desire to keep each period relatively independent in an accounting sense, the difficulty of asking subjects to pay the experimenter, and our desire to assure subjects their show-up fees. As with the previous constraint, we also expected that our rule preventing received punishment from pushing earnings into the negative range it would rarely be activated. In practice, punishment received exceeded first stage earnings in only 23 of 2960 subject-periods (0.78%). Finally, we chose to have no exception to incurring the cost of punishing because costliness to the punisher is too important a feature to waive, in the sense that the standard prediction of zero punishment requires it. In the event, only 5 of the 2960 subject-periods (0.17%) under IS saw a subject incur negative earnings for the period, and no subject's aggregate earnings from the session's main experiment approached 0 points (the lowest observed being 437 points (\$12.85 before show-up fee), occurring in the exogenous IS comparison treatment).

$$\max\left\{(20 - C_i) + 0.4 \cdot \sum_{\text{All } j \text{ in a group}} C_j - 4 \cdot \sum_{\text{all } j \neq i} p_{ji}, 0\right\} - \sum_{\text{all } j \neq i} p_{ij} \quad (2)$$

which differs from Eq. (1) of Section 2 by substitution of specific values for E_i , m and P and by the imposition of the floor on earnings other than from cost of punishing indicated by the max function.¹²

When FS is in place, each period begins with voting on which of the two allocation possibilities is subject to sanctioning, reporting of the majority vote outcome to group members, and choice of preferred sanction rates from the set 0, 0.4, 0.8 and 1.2 points. For each point a group member allocated to the account in question, he loses an amount equal to the chosen sanction rate. The median preferred rate in the group is taken as the group's choice and announced to its members.¹³ Subjects then make their contribution decisions, are shown one another's contributions in a random order, and learn their earnings for the period before repeating the process again. Note that a subject's privately optimal strategy remains zero contribution if a rate of 0 or 0.4 is set but is contributing her full endowment if the rate is set at 0.8 or 1.2 points (see equation (3)). If a group member is sanctioned, one third of the amount deducted from that member's earnings is also charged to the group as a whole as a variable cost of sanction administration and is divided equally among the five members. Thus each member including the punished individual bears a cost equaling one fifteenth of the sanction itself, causing the ratio of cost collectively born by the four other members to total cost to the sanctioned individual to be 1:4, as in the IS treatment.¹⁴ Also as in the IS treatment, the cost of imposing sanctions, including the fixed cost, is always born by each individual even if it makes her earnings for the period negative, whereas the effective sanction itself cannot exceed the individual's first stage earnings for the period. Thus, an individual's earnings under FS are given by

$$\max\left\{(20 - C_i) + 0.4 \cdot \sum_{\text{all } j \text{ in a group}} C_j - p_i, 0\right\} - \frac{1}{15} \sum_{\text{all } j \text{ in a group}} p_j - f \quad (3)$$

¹² Adding the earnings floor is not expected to alter behavior sufficiently to require us to revisit our theoretical predictions.

¹³ The instructions simply state that "the fine rate selected will be the median of the preferred levels entered by individual group members."

¹⁴ We made the ratio the same for both kinds of sanctions and referred to the uniformity of 1:4 ratio under the two schemes in the instructions so that the difference in variable cost per sanction would not in itself influence subjects' votes. Realized variable sanction cost could still be lower in FS if fewer sanctions are needed due to the certainty with which they were expected, as we will show in fact to have been the case.

where $p_i = r \cdot (20 - C_i)$ is the imposed sanction if contributing to the private account is penalized,¹⁵ with $r \in \{0.0, 0.4, 0.8, 1.2\}$ being the sanction rate, and $f \in \{0, 5\}$, depending on the treatment, is the fixed or administrative cost of FS. As should be clear from (2) and (3), the costs of a sanction scheme and of any fines paid are lost both to individual group members and to the group as a whole and do not, e.g., serve to fund provision of the public good.

Insofar as possible, instructions are presented in natural language comprehensible by those without a strongly mathematical orientation but avoiding terms that might suggest that contributing to the group account is good or that those contributing less to that account should be punished. We use “allocate” rather than “contribute,” “assign reduction points” rather than “punish.” The term “sanction” is not used, and although “fine” is used, we even-handedly indicate that this can be a fine for allocating points to the private account or for allocating point to the group account. The institutions voted on are called “individual reduction decisions” (= IS) and “group-determined fines” (= FS), and the fixed cost of FS, in the treatments including it, is called a “fixed administrative cost of having a fine scheme in operation.”

All treatments also include three additional tasks, two of which have payoff consequences. First, before subjects make their first contribution decision they fill out a table of the type used in Fischbacher *et al.* (2001) and Fischbacher and Gächter (2010), indicating a contribution choice conditional on the average contribution of others, which has 21 possibilities given rounding to the nearest integer. The payoff consequence occurs in an “additional period” following the first ordinary (unconditional) contribution decision.¹⁶ In that period, one group member is randomly chosen, and contributions for the period consist of the chosen individual’s conditional contribution under the other members’ initial unconditional contribution choices, and of those four initial contributions, with the earnings

¹⁵ $p_i = r \cdot C_i$ if contributing to a group account is penalized.

¹⁶ In the 6-Vote treatments, the outcome of the additional period is shown to subjects after the one regular contribution period (under NS condition) in Phase 0. In the 3-Vote treatments, this occurs after the first of the four regular contribution periods in Phase 1.

of each determined accordingly.¹⁷ We included this task in order to have a measure of cooperative orientation when analyzing subject voting and contribution decisions.¹⁸

Second, after the main experiment ends, subjects take an intelligence test consisting of 15 problems. Subjects earn 15 points for a correct response and lose 3 points for an incorrect response, with average point earnings being less than 1/10 those of the main experiment. The questions are taken from Séréville and Myers (1994) and are based on Raven's progressive matrices (see the Appendix for an example). As an indicator of intelligence, we use the number of correct responses. We include these questions to test whether voting is affected by cognitive ability. The scores, which we refer to as IQ, also allow us to check whether contributing to the public good and voting are affected by intelligence.

Third, subjects are asked a set of questions about their political views which we use to test for effects of such views on choice of institutions in the main experiment, as in Putterman *et al.* (2010). Since there are no correct answers, this portion has no payoff consequences, although all subjects are required to complete it. In the initial instructions, subjects were told that the session included a main experiment and two shorter portions (the conditional contribution schedule and "additional period" being treated as part of the main experiment). Each session ended with a few brief questions about gender, major, number of economics courses taken, and a request for open-ended comments on the experiment. The time-lines of the BASELINE, 3-Vote and 6-Vote treatments are shown in Figure 1. All instructions are included in an online Appendix.

¹⁷ The instructions indicate that the conditional decisions will affect earnings in that one period only, and the additional period is neither more heavily nor less heavily weighted than the other 24 periods, from a payoff standpoint.

¹⁸ A possible concern is that elicitation of contribution choices conditional on others' choices has a framing effect on subsequent behavior. Comparison with behaviors in Markussen *et al.*, 2011, in which there is no conditional contribution task, provides no evidence that the task alters unconditional contribution decisions.

4. Results

4.1. The experiment

14 sessions, two for each treatment, were conducted in a computer lab at Brown University during the fall of 2009. We conducted two additional sessions of the exogenous IS comparison treatment (discussed below) with 30 subjects in spring of 2010 resulting in a total of 300 participants. Subjects were Brown undergraduates in a broad range of fields, recruited by e-mail solicitation of registrants of the BUSSEL recruiting system, with no subject participating in more than one session.¹⁹ An experimenter read all instructions aloud as participants read along, and subjects answered control questions which were included after each set of instructions to test comprehension and asked any questions of their own before the decision portion(s) began.²⁰ The experiment was programmed in Z-Tree (Fischbacher, 2007). From signing of consent forms to completion of debriefing questions, sessions took slightly under two hours. Subjects were paid privately in cash at the end of their session, with earnings including a \$5 show-up fee, an average of \$2.22 for the IQ questions and an average of \$24.35 for the main portion of the experiment.²¹

4.2 Voting for formal vs. informal sanctions

Table 2 summarizes the individual votes and majority decisions between formal and informal sanctions in the 93 votes of the 3(FI)-N, 3(IF)-N, 3(FI)-C and 3(IF)-C treatments and in the 90 votes of the 6-N and 6-C treatments, and Figure 2(a) shows the pattern of change in voting outcomes over time, with the data grouped into 3-Vote and 6-Vote

¹⁹ Almost all subjects listed multiple concentrations (majors), with 64 out of 300 subjects (23%) having economics as one of them, about twice the share of economics concentrators in the university's undergraduate population as a whole. Exactly half of the participants were female, similar to the undergraduate gender ratio at Brown. BUSSEL (Brown University Social Science Experiment Lab) uses software based on CasselWeb2. Students register in response to flyers and campus newspaper ads (both addressed to all students, without reference to major) referring to the possibility of earning money by participating in decision-making experiments run by researchers in the Department of Economics. Students who took part in previous voluntary contribution experiments at the university were barred from participating.

²⁰ In 3 Vote treatments, questions were taken and answered after the reading of instructions before each of phases 1, 2, 3 and 4. In 6-Vote treatments, questions were taken and answered after the reading of instructions before phases 0 and 1, and in BASELINE, after the reading of the only instructions, those before phase 1. For the full instructions, see the Appendix.

²¹ Average earnings for the main portion of experiments are \$18.02 in the BASELINE treatment, \$25.47 in the 3(FI)-N and 3(IF)-N treatments, \$24.52 in the 3(FI)-C and 3(IF)-C treatments, \$27.90 in the 6-N treatment, and \$25.10 in the 6-C treatment.

treatments and treatments with and without administrative cost in the FS scheme. The first result that stands out is that contrary to the prediction of standard theory that subjects would uniformly select formal over informal sanctions, the choice between the two schemes only slightly favors formal sanctions (102 group votes) over informal ones (81 group votes), with the administrative cost of FS serving as a major deterrent to that system's use. Whereas 75 out of 87 group votes (86%) in the no administrative cost treatments favored FS over IS, the corresponding numbers were 27 out of 96 votes (28%) in treatments with 5 point administrative cost.

The order in which subjects were exposed to the FS and IS conditions in the 3-Vote treatments appears to have had no significant effect—as shown more formally below. Whether subjects were first exposed to FS and IS conditions exogenously and then voted (3-Vote treatments), versus voting from the outset (6-Vote treatments), also had little qualitative impact on vote outcomes, although we observe a smaller margin of preference for IS over FS in phases 4 – 6 of the 6-Vote than of the 3-Vote treatments with administrative cost, perhaps indicating a slight shift towards FS with learning. We explore the determinants of individuals' votes in Section 4.6 after first reporting how the performances of the competing institutions compared.

4.3 Voting on formal scheme parameters

The 102 group votes in favor of formal sanctions make it the case that there were 408 periods in which groups voted on the parameters of a formal sanction scheme, with 2,040 individual votes on whether to penalize contributions to the private or public account and the same number of individual votes for penalty rate 0.0, 0.4, 0.8 or 1.2 per point assigned to the account selected. There was also voting on the parameters of FS in periods when it was imposed exogenously, accounting for another 124 periods of group voting on those parameters, or 620 individual votes, during periods 5 – 8 of the 3(FI)-N and 3(FI)-C treatments and periods 9 – 12 of the 3(IF)-N and 3(IF)-C treatments. The preponderance of outcomes—520 of 532 group votes (97.7%)—favored sanctioning contributions to the private rather than public accounts, with 39 of the 42 groups that used FS in at least one phase voting exclusively for this efficient choice (see Appendix Table B1). Individual votes

to sanction contributions to the public accounts were slightly more common than were corresponding majority outcomes at the group level, with 64 of the 210 subjects (30.5%) who had at least four opportunities to vote on the issue voting one or more times to penalize allocations to the public accounts. Altogether, 355 of 2,660 individual votes cast (13.3 %) were for this inefficient option.²²

In the 520 group votes for penalty rates under the rule of penalizing allocations to the private accounts, 433 (83.3%) chose penalty rate 1.2 and 34 (6.5%) chose rate 0.8, so that in almost ninety percent of cases in which groups voted to penalize contributions to the private accounts, they put a binding sanction in place. The non-binding 0.4 rate was selected in 34 cases, and the rate 0 in 19 of the cases in which contributions to the private accounts were selected for penalty. However, in every period in which a group had chosen to penalize contributions to their public account, they set the penalty rate at 0.²³ Of the individual votes under formal sanctions for contributing to the private accounts, 71.3% were for the 1.2 rate, 5.3% for 0.8, 4.5% for 0.4 and 18.9 % for 0.0.

Figure 2(b) shows the proportion of median votes cast for each sanction rate including sanctioning of public account contributions, which is observed only at the rate 0. One interesting observation is that while choosing to penalize contributions to private accounts at the maximum rate of 1.2 is the preponderant choice (89.2% of votes) in the treatments without administrative cost, its dominance is less overwhelming (65.1% of votes) in the treatments with administrative cost, perhaps because subjects were reluctant to add to the fixed cost of 5 points per period that they already bore when using FS (a sunk cost fallacy?). There are also indications of a learning process, with the share of (FS-using) groups fining contributions to private accounts at the maximum rate rising from less than half in Phase 1 to 100% in Phase 6 of the 6-vote treatments.

²² We report separately group and individual votes under exogenous versus endogenously chosen formal sanctions regimes in Appendix B. The much higher proportion of subjects who voted at least once to penalize contributions to the group account than of total votes for that choice is consistent with the possibility that many votes of this kind were cast by an individual just once and in error. However, the proportion of individual votes in favor of penalizing contributions to the group account shows only a mild downward trend over time in the 6-Vote treatments, with such a trend present if at all in the 3-Vote treatments only when periods in which FS as such is imposed exogenously (periods 5 – 12) are excluded.

²³ A plausible interpretation is that subjects voted to penalize contributions to the public rather than the group account when their objective was to avoid penalties altogether.

4.4 Contribution levels and trends by condition

The left panel of Figure 3 displays the average contribution to the public account by period and condition, distinguishing between formal and informal sanctions schemes when imposed exogenously versus when adopted by vote. The dashed curve shows the contribution trend in the BASELINE treatment and in Phase 1 of the 3-Vote treatments, which follows the familiar VCM pattern of beginning around 50% of endowment and declining with repetition.²⁴ In contrast, groups interacting under either sanctions regime, whether with exogenous or endogenous regime choice, on average exhibit strong upward trends in contributions in Phases 1 – 3 and stable and high contributions above 90% of endowment in Phases 4 – 6.

In all phases, regardless of order of introducing IS and FS, presence or absence of administrative cost, and whether a scheme is exogenous or endogenous, contributions are significantly higher with either IS or FS than in the NS regime in Phase 1 of the 3-Vote treatments and all phases of the BASELINE treatment.²⁵ Comparing the impacts of informal and formal sanctions, we find that with both voted and exogenous sanctions, average contributions are higher in the first three phases of using sanctions when FS is used than when IS is used. This difference is often statistically significant, although it is economically small in comparison to the difference between either FS or IS, on the one hand, and NS on the other hand (see Figure 2). Indeed, in the last three phases, contributions do not significantly differ between FS and IS. In summary, then, both IS and FS raise contributions relative to NS, with an initial advantage for FS but full catching up by IS with experience.

²⁴ The fact that contributions fluctuate without trend during periods 5 – 18 is also not unusual given the partner matching protocol in which there may be attempts to restart cooperation, possibly facilitated by the pause in play following each set of four periods (see again Andreoni, 1988).

²⁵ According to group-level Mann-Whitney tests, in all phases, the average contributions of groups playing under exogenous FS, exogenous IS, endogenous FS or endogenous IS are significantly higher than those of groups in BASELINE treatment, which use no scheme (NS). In almost all cases, the significance level is 1% level (see Appendix Table B.2). Although the number of observations becomes small, even if we compare the average contributions with endogenous FS only to those with no scheme, or if we compare those with endogenous IS only to those with no scheme, in all phases except the cases in which the number of observations is 1, the difference in contributions is statistically significant.

That contributions to the public accounts are high when groups impose sanctions of 0.8 or 1.2 on contributing to the private accounts is unsurprising. An interesting question to ask is whether non-deterrent sanctions at rates of 0.4 or 0.0 affect contributions. Figure 4 shows average contribution by phase under formal sanctions, disaggregated by sanctions rate but pooling observations from all treatments, including exogenous formal sanctions in Phases 2 and 3 of the 3-Vote treatments. For reference, average contribution under NS regimes (Phase 1 of the 3-Vote treatments and Phases 1 – 6 of the BASELINE treatment) is indicated by the dots in the middle of each set of bars (which we join by dashed line for clarity and to show the contribution trend under NS). The data show that subjects contribute substantially more at sanction rate 0.4 than in the absence of sanctions, a finding that is consistent with past evidence of sensitivity of contributions to the MPCR but that is also likely to reflect a signalling effect, as with the non-deterrent sanctions in Tyran and Feld (2006).²⁶ Although average contributions are higher at sanction rate 0 of the FS regime than in the NS regime in five out of six phases, the difference is generally small, and there are too few group level observations to test for significance.²⁷

4.5 Punishment under informal sanctions condition

The finding that contributions increase and stabilize at high levels in the face of the opportunity to impose informal sanctions suggests that decentralized costly punishment of low contributors was not rare in this condition. The right or (b) panel of Figure 3 shows,

²⁶ The result is comparable to past findings that subjects contribute more when the MPCR is higher, since when a penalty rate of 0.4 is in place, an individual gains 0.2 from allocating a point to her private account, comparable to facing an MPCR of 0.8, whereas without penalties the individual gains 0.6, since the MPCR is 0.4. An important difference between cases of non-deterrent sanctions in the present design and that in Tyran and Feld (2006) is that subjects in that paper's non-deterrent sanction treatment did not have the option of voting for a stronger sanction level, so their votes might be expected to have sent a clearer signal of desire to cooperate than is the case with our design. The fact that contributions were nonetheless so high at the 0.4 rate especially in Phases 1 – 4 of our experiment suggests that many subjects interpreted their group's votes for formal sanctions, for penalizing contributions to the private accounts, and for a positive penalty rate as signaling desires to coordinate on cooperative actions. Conceivably selection of 0.4 rather than 0.8 or 1.2 was viewed by some as an indication that group members preferred a gentle nudge to a harsh penalty. Markussen *et al.* (2011) offer only non-deterrent sanctions in half of their treatments and find, similar to the present paper, that they significantly increase contributions, although less so than the deterrent sanctions that are made available in others of their treatments.

²⁷ Mann-Whitney tests using average contributions of each individual are not strictly speaking valid due to the non-independence of observations. If we conduct such tests anyway simply to obtain an informal impression, we find that contributions under formal sanctions with rate 0 and those in BASELINE treatment appear to differ only in Phase 5 and only at the 10% level.

separately for endogenous and exogenous IS, the average amount of informal sanctions *given*, on a per subject basis (sanctions *received* cost those targeted four times the numbers of points shown). For both endogenous and exogenous IS, sanctions begin at around 1.6 points per subject and decline with repetition, reaching about 0.3 or 0.4 points per subject in later periods of endogenous IS. The figure also displays the amount of perverse punishment during the period in question, demonstrating that punishing contributors of more than the group median was not unknown but that with the exception of one period, it accounts for a small fraction of the total (overall, 5.53% of punishment points go to above-median contributors). In all, 75.3% of those subjects who play at least one phase in IS condition engage in costly punishment at least once, with 68% (43%) engaging in costly punishment in more than one out of four (one out of two) of those occasions in which group members contributed unequal amounts, and the opportunity to punish was available. The uptick in punishment in period 24 confirms that punishment is not entirely strategic in nature (compare Falk, Fehr and Fischbacher, 2005), although a few subjects appear to have gambled that others would not bother to punish in the last period (see panel (a)).

Evidence that punishment has a direct bearing on contributions can be obtained by estimating regressions in which the change in an individual's contribution from period t to period $t + 1$ is the dependent variable, and the punishment points deducted from his or her earnings in period t is an explanatory variable. Past studies, such as Önes and Putterman (2007), suggest that subjects respond differently to being sanctioned depending on whether their contribution was low or high compared to others in their group at the time; we accordingly include interaction terms between punishment received and dummy variables for being an above-median or a below-median contributor, with median contributor the omitted category. We also include stand-alone above- and below-median contributor dummy variables because subjects may adjust their contributions independently of whether they receive punishment, and we include a possible trend term and subject fixed effects. These regressions, shown in Appendix Table B.3, strongly support the proposition that punishment of low contributors leads them to increase their contributions, while punishment of high contributors has a mildly contrary effect, justifying its description as perverse. Interestingly, below- (above-) median contributors also tend to increase (decrease) their

contributions independent of receiving punishment, perhaps both because of general conformism and because the pattern of punishment makes median or average contributors least likely to attract it. (Details are shown in Appendix Table B.3)

4.6 Effect of sanctions on earnings

The social dilemma faced by our subjects is illustrated by the fact that each can earn 40 points per period if all contribute their full endowments to the public accounts versus 20 if all fully free ride; but free-riding is individually rational since an individual always earns more when contributing less, taking others' contributions as given. The possibility of sanctions can resolve this dilemma, but does so most efficiently when sanctioning costs are avoided, i.e. when the expectation of sanctions suffices to deter free-riding. Maximum average earnings of 40 per period would be achieved in any period of IS or of FS without administrative cost (35 per period in FS with administrative cost), if full cooperation were to be achieved without actual imposition of sanctions. Because no group exhibits 100% free-riding when in NS condition, however, it is good to keep in mind achieved efficiencies in that treatment for purposes of comparison. In the BASELINE treatment, earnings averaged 26.8 in Phase 1 and 24.7 overall, per subject and period. 3-Vote Treatment subjects earned an average of 28.8 per subject and period under NS in Phase 1, with 6-Vote treatment subjects earning an average of 27.7 in their one NS period 0.

Figure 5 shows average earnings in early and late phases of the experiment under FS, IS, and NS conditions, with panels (a) and (b) showing the data for 3-Vote treatments respectively without and with administrative cost, and panels (c) and (d) the data for 6-Vote treatments respectively without and with administrative cost. For purposes of comparison, average earnings in the BASELINE treatment are also shown (identically) in each panel. (The trends of average earnings are graphed by period in Figure B.3 of the Appendix.) We group the early and late phases to check for impacts of experience or learning, and because in the 3-Vote treatment the IS and FS conditions are experienced only exogenously in phases 2 and 3, versus endogenously in phases 4 – 6. The figures show that the earnings achieved under IS were appreciably higher in the later than in the earlier phases in both 3- and 6-vote treatments, while earnings under FS show smaller gains, mainly in the 6-Vote

treatments. The change in earnings over time is statistically significant for the IS but not the FS condition.²⁸ As for the comparison of IS and FS earnings during given phases, the figure shows higher average earnings under FS than under IS in the early periods, differences that are significant only in phases 2 and 3 of the 3-vote treatments without administrative cost.²⁹ Apart from the latter treatments, in which the ordering remains unchanged, for phases 4 – 6 average earnings under IS exceed those with FS, with the difference being statistically significant in the 3-vote treatments with administrative cost.³⁰ We conclude that a fixed cost of publicly administered sanctions as small as 5/40 or 12.5% of potential earnings (or 25% of the gain from shifting from zero to full cooperation) is sufficient to render the informal sanctions regime as efficient as or more efficient than the formal one following initial experience. Even when there is no administrative cost, IS achieves similar efficiency to FS after experience, although it is not chosen often enough for statistical testing. These results contrast with standard theory's prediction that informal punishment opportunities should leave contributions at the own-payoff maximizing level of zero and that, accordingly, formal sanctions costing less than 20 points per period should always be opted for.³¹

²⁸ For group-level earnings in IS phase 2 or 3 and average earnings in the same groups when using IS in phases 4 – 6, Wilcoxon Signed Rank tests show the difference in the 3-vote treatments without (with) administrative cost to be significant at the 5% (1%) level. Parallel tests find no difference in earnings of phase 2 or 3 versus phases 4 – 6 for FS in either set of 3-vote treatments.

²⁹ For the latter treatments, $p = .002$ in phase 2 and $p = .036$ in phase 3, in 2-tailed Mann-Whitney tests. That test can't be carried out for the **6-N** treatment because only one group voted for IS.

³⁰ For the 3- and 6-vote treatments without administrative cost, tests cannot be performed because only one group in each case voted to use IS. For the 3-vote with administrative cost, only phase 4 has sufficient groups voting for FS, so the Mann-Whitney test for earnings difference, which gives p -value of 0.033, is for phase 4 only. In **6-C** treatment, two groups continuously voted for FS and four for IS, so a Mann-Whitney test can be performed for phases 4 -6, yielding a p -value of 0.340.

³¹ Since the reported tests for difference of earnings under IS versus FS leave out some observations—e.g., those of groups using only one scheme or the other in the early or late part of their session—we thought readers might like to know overall average earnings under each scheme in each treatment or treatment pair, weighting equally all periods in which a scheme was used, whether exogenously or endogenously, early or late. These grand average earnings under IS and FS respectively are 33.13 and 38.10 in the 3-vote treatments without administrative cost, 36.18 and 33.09 in the 3-vote treatments with administrative cost, 37.43 and 37.96 in the **6-N** treatment and 33.98 and 33.25 in the **6-C** treatment. Thus, average earnings are about 5 points higher per period under FS than under IS in the 3-vote treatments without administrative cost and about 3 points higher per period under IS than under FS in the 3-vote treatments with administrative cost, and average earnings under each scheme are quite similar in the two 6-vote treatments. Average earnings with IS are lower in those treatments in which IS is used little partly due to the lessened benefit of experience in those treatments. Interestingly, the difference in average earnings under FS in those treatments with versus those without administrative cost is close to 5 points, i.e. the cost itself. Although an appropriate formal test using all of the data is difficult to identify, one way of using most of the data is to take, for each group that used both IS and FS at least once during their session, average earnings under IS and average earnings and under FS (disregarding differences in the number of phases entering each average) and perform a Wilcoxon test. For both sets of 3-vote treatments paired in Figure 5 and in the above discussion, differences between IS and FS

An observed feature of IS experiments is that efficiency may increase over time if contributions rise while the level of active sanctioning falls (see Gächter, Renner and Sefton, 2008). This pattern is observable for our IS-condition subjects in Figure 5. On average, subjects in our experiment earned 26.3 points per period during the first phase in which their group used IS, versus 37.1 in the last such phase (this calculation includes only groups using IS in at least two phases). The difference in initial versus final phase earnings with IS is statistically significant at the 0.05 level according to a Wilcoxon Signed Rank test.³²

An interesting question regarding informal sanctions is whether the amount given is sufficient to cause contributions to increase by reason of self-interest—that is, are earnings actually higher for those members of a group who contribute more, due to the presence of punishment? To investigate this, we estimate, using all observed interactions under the IS condition, subject-level regressions with individual fixed effects and robust standard errors clustered by group. The dependent variable is earnings and the independent variables are individual contribution and a constant. We find (see Appendix Table B.6, part (1)) that contribution is a significant positive predictor of earnings, meaning that punishment did in fact reverse the incentive to free ride.³³

earnings are statistically significant according to this test, with $p = .003$ in the case of 3-vote treatments without administrative cost and $p = .034$ in the case of 3-vote treatments with administrative cost. The same kind of test cannot be performed for treatment **6-N** observations because only 1 group undertook play under both schemes. The test can be performed for **6-C** and yields $p = 0.650$.

³² We also test for a trend in earnings under IS by estimating a linear regression whose dependent variable is average earnings per group per phase using IS, and whose independent variable is the number of phases the group has used IS thus far $\{=1, 2, \dots\}$. We include group fixed effects. The number of phases the group has used IS obtains a positive significant coefficient of 1.75 which is significant at the 1% level.

³³ While demonstrating that contributing more raised earnings on average, the regression specification using a single measure of own contribution cannot rule out that it is most profitable to contribute only a little more than the group average, with earnings not rising further for additional contributions beyond that. We tested for this possibility by estimating alternative specifications of the above regressions in which we substituted for the contribution variable (i) the absolute negative difference between own contribution and group median contribution ($= 0$ if $C_i \geq C_{med}$), and (ii) the positive difference between own contribution and group median contribution ($= 0$ if $C_i \leq C_{med}$). In these regressions (see Appendix Table B.6, part (2)), the negative deviation term obtains a significant negative coefficient, meaning that the further one's contribution is below the median, the less one earns, consistent with the general result above. But the positive deviation term also obtains a negative coefficient, meaning that contributing more fails to raise earnings for contributions above the median. This is consistent with the typical finding that punishment increases with distance of contribution below the average or median but is relatively unrelated to contribution at higher levels, possibly even increasing somewhat for those contributing the highest observed amounts (the perverse punishment phenomenon).

4.7 Explaining individuals' votes

Figure 5 shows that in the second half of each session, whether earnings were higher under IS or FS depends largely on the absence or presence of the administrative cost in the FS condition (the near tie in 6-N treatment is a partial exception). The preponderance of voting for FS in the C and for IS in the N treatments is broadly rationalizable by these patterns. However, apart from the 3-Vote treatments without fixed cost, in which earnings were higher under formal than under informal sanctions in both the early and late parts of sessions, neither sanction scheme dominates the other in efficiency at all times. Since group results varied and since subjects received no feedback about other groups' outcomes, we study the impact of earnings on voting between schemes with a closer eye by estimating a series of probit regressions, the results of which are displayed in Table 3.

The regressions use observations for the votes of Phases 4 – 6 of the 3-vote treatments because relative earnings under both schemes is always available, thanks to the exogenous phases all subjects experience during the earlier phases in those treatments.³⁴ We pool observations across exogenous condition orders FI and IF, but we separate, into Tables 3.a and 3.b respectively, the observations of treatments without and with administrative cost, identified by N and C respectively. The column (1) regressions show that subjects' relative earnings influenced their votes in an intuitively sensible manner, statistically significant at the 1% or 5% levels respectively for the C and N treatment observations. The coefficient on the Phase 2 dummy suggest that whether subjects were in treatment 3(FI)-N or 3(IF)-N does not affect voting. The column (2) regressions add the ratio of experienced coefficients of variation of earnings under the two schemes to check whether variability of earnings matters; it appears that it does not.

The regressions of columns (3) and (5) add as controls binary indicators of whether a subject punished a contributor of more than their group's median contribution or received punishment when a contributor of more than the median—events we call “perverse

³⁴ Not only is it relatively rare for groups to have switched regimes in the 6-Vote treatments, but the fact that the order in which they did so depends upon their votes introduces an element of endogeneity that would be difficult to control for.

punishment.”³⁵ Perverse punishment is singled out because this “misdirected” informal sanction has been shown in other studies to be a major dampener of efficiency.³⁶ The regressions of columns (4) and (5) add controls for three personal characteristics: average conditional contribution in the strategy task, taken as a possible indicator of cooperative orientation; IQ; and gender [female = 1].³⁷ Two of the added coefficients attain significant coefficients. First, individuals who perversely punished high contributors are significantly less likely to favor a formal sanctions regime in the treatments without administrative cost (but the variable shows no significant effect in the treatments with such cost). Perverse punishers thus seem to show their distaste for cooperation by opposing a sanctions regime that would effectively force group members to contribute, manifesting this preference in precisely those treatments in which most group members favor such sanctions. Second, IQ appears to have a significant effect in both treatments, but of opposite sign. In particular, higher IQ subjects are more likely to favor formal sanctions in treatments without administrative cost and more likely to favor informal sanctions in treatments with such cost. Thus, higher IQ is associated with favoring precisely that scheme which the majority as a whole favors and that proves most efficient *ex post* in the no cost and administrative cost treatments respectively—a result that seems consistent with the strong cognitive skills they display in the IQ task. The result equivalently implies the intuitive idea that low IQ individuals were more likely than others to have made mistakes when casting their votes for one scheme versus the other.

³⁵ The relevant punishment event or events need to have occurred during the periods of exogenous IS, that is periods 5 – 8 for 3(IF)-C and 3(IF)-N treatment subjects, and periods 9 – 12 for 3(FI)-C and 3(FI)-N treatment subjects.

³⁶ Calculations in Cinyabuguma, Page and Putterman (2004) show that presence of perverse punishment can account for failure of informal sanctions to raise efficiency. Ertan, Page and Putterman (2009) find that making informal sanctions available fails to raise efficiency until perverse punishment is ruled out. Önes and Putterman (2009) show that high contributors respond to being punished by reducing their contribution, a result which we reconfirm below. Whereas those authors define punishment as perverse when the recipient’s contribution is above the group average for the period, we take advantage of our five rather than four member group size to adopt a slightly stronger criterion. Cases of perverse punishment overlap, in the large majority of cases, the phenomenon labeled “anti-social punishment” by Herrmann, Thöni and Gächter (2008), who provide strong evidence of its negative impact.

³⁷ Average conditional contribution is simply the average of the individual’s 21 entries in the form indicating what he or she chooses to contribute assuming others on average contribute 0, 1, ..., 20. While Fischbacher *et al.* (2001) classify subjects as conditional cooperators, free rider, etc., in Putterman, Tyran and Kamei (2010) we find average conditional contribution to be as good an indicator of conditional willingness to cooperate as individual type dummies or other statistics based on the conditional contribution schedule.

4.8 Did endogenous choice increase the efficiency of sanctions?

Because the levels of contributions and efficiency observed under endogenous IS regimes in the 6-Vote treatments and in phases 4 – 6 of the 3-Vote treatments are on average quite high, it seemed possible to us that part of the efficiency of voted IS could be due to a signaling or other effect of having being democratically chosen. To investigate this, we conducted two sessions, each with three groups, collecting six observations of a treatment in which groups operated under the IS condition as a result of its exogenous imposition, rather than of voting. The test treatment was structured identically to treatment 6-C³⁸ except that in the instructions before Phase 1, when the IS and FS conditions were explained to subjects, they were told that which condition they would play each phase under would be determined by the computer, with no mention being made of the possibility of voting.³⁹ In fact, the computer program assigned the subjects in all groups of this treatment to the IS condition in every phase, so that six group-level observations could be collected paralleling the groups in the 6-C treatment that voted for, and thus played under, the IS condition in each of phases 1– 6.

Figure 6 compares contributions and earnings in treatment 6-C groups that voted to use IS to those in groups in the exogenous IS test treatment. For treatment 6-C, we plot both the average contribution and earning level for members of groups that have only used IS up to and including the phase in question, and the average contribution and earning level for members of the four groups that use IS in all six phases of their session.⁴⁰ Average

³⁸ We chose treatment 6-C since it exhibits the most play under voted IS, thus providing more comparison groups.

³⁹ Subjects in the test treatment had not participated in any other treatment and with high likelihood had no knowledge that similar experiments had been conducted which included voting. The instruction regarding the determination of whether IS or FS would be used read as follows: At the beginning of the first period in every 4 period phase, you will be informed which of two schemes, “Group-determined fines” or “Individual reduction decisions,” will be in effect in your group in that phase. (There will be a pause lasting 40 seconds in between each pair of phases, and the announcement will come at the end of that pause.) You may be using a different scheme in a phase than was used in the previous one, or the same scheme in several or all phases. We cannot tell you on what basis the scheme you are to use will be decided, but we can assure you that which scheme you will use in a given phase is unaffected by your own and your group’s behaviors during the previous phase or phases.

⁴⁰ A group that votes for IS in an early phase or phases but subsequently votes for FS is included in the data plotted in the first curve but not the second. Data for a group that votes for IS in any phase coming after one in which it voted for and used FS is included in neither curve. The number of groups included in the first curve is 6, that in the second 4. (Table 2 (a) shows that the number of groups voting for IS returned to 6 in phase 3, then settled at 5 in phases 4 – 6.)

contribution and earnings rise over time as shown by all curves, with both contributions and earnings in the endogenous IS groups tending to exceed those in the exogenous IS groups by roughly four points. The visual impression from the figure is therefore quite consistent with the conjecture that voted choice is associated with better performance of the IS scheme. However, the appropriate formal tests are non-parametric tests using group level observations, of which the available number is small. It turns out that with some performance variation across groups, the observed differences are not statistically significant, or are at best weakly so—i.e., at the 10% level using a one-tailed test.⁴¹

5. Discussion and Conclusion

If all individuals were fully rational, strictly self-interested, and had common knowledge of this, then (unless the conditions of an infinitely repeated game are present and a cooperative equilibrium obtains) there should be no contribution to a public good if the private marginal return is less than the endowment good's private value. Members of a group that had access to a formal sanctioning mechanism—one that it is natural to identify as a “state” in problems of societal scale—should accordingly favor use of that mechanism if its per person cost to deliver perfect cooperation is less than the individual gain from shifting from no to full cooperation. When such preferences and common knowledge of them are assumed, voluntary cooperation is in theory possible neither without nor with opportunities for individuals to impose costly penalties on one another, so it is recourse to the mechanism of formal sanctions that should always be observed.

We conducted an experiment in which subjects in fixed groups of five anonymous partners played a finitely repeated linear voluntary contribution game wherein the private return from contributing to the public good is too low to warrant contributions by selfish agents with common knowledge, but in which full contribution of endowments is socially optimal. In this setting, we offered our subjects the opportunity to elect to use a formal

⁴¹ The Mann-Whitney tests at group level are reported in Appendix Table B.8. With respect to contributions, the table shows *p*-values for two-tailed tests that range between 0.10 and 0.20 in Phases 1, 4 and 5; the one-tailed test *p*-value is exactly half that of the two-tailed test, hence the difference is significant at the 10% level for these three of six voting phases. For average earnings, only Phase 1 meets this criterion, and only when the six groups that selected IS in Phase 1, including two that went on to select FS in Phase 2, are included.

sanction mechanism, the parameters of which they could set by vote so as to render full cooperation the unique equilibrium, or instead to use an informal sanction mechanism. Under the latter, each individual could make free riding costly to others, but only at a cost to herself or himself such that zero punishment and zero contributions is the only equilibrium under the rationality, self-interest, and common knowledge assumptions. We varied across treatments whether subjects encountered the formal and informal sanction regimes exogenously before voting on them or began voting without prior experience, and we also varied the shared administrative cost of using the formal sanctions mechanism, setting it at zero in half of our treatments and at a quarter of the potential gain from cooperation, in the other half.

Unlike the theoretical prediction, the large majority of subjects voted to use the formal sanctions mechanism only when it carried no fixed cost, with the majority instead selecting the informal sanctions regime when formal sanctions were costly. A major reason for this, undoubtedly, is that as in the other experiments reported in the past decade, subjects succeeded in reaching high levels of efficiency even with informal sanctions. Those sanctions were used sufficiently often and in a sufficiently well-targeted manner that in their presence, it became payoff-maximizing to contribute at least the observed group average level to the public good. Indeed, even when formal sanctions carried no administrative cost, some groups chose the informal sanctions mechanism and achieved contribution and earning levels indistinguishable from those of groups adopting formal sanctions. While the relatively high efficiency of play under the informal sanction regime in our experiment as compared to other experiments may in part be attributable to the high effectiveness of sanctions (a 1:4 ratio of cost to punisher versus cost to recipient of punishment—see Nikiforakis and Normann, 2008) and to a subject pool favorably disposed to norms of cooperation (see Herrmann *et al.*, 2008), by conducting an additional treatment in which the informal sanctions scheme was assigned to groups exogenously, we also found indications that being endogenously chosen by the subject groups played a part in the mechanism's efficient use.

Our paper shares the aim of studying subject choice over formal and informal sanction mechanisms with a companion paper, Markussen *et al.* (2011). Their experimental

treatments (implemented using subjects at the University of Copenhagen versus the present paper's subject pool at Brown University in the U.S.) share core features with those studied by us including group size, MPCR, partner grouping, finite repetition, punishment cost, and effectiveness under the informal sanctions scheme, and additional treatments using exogenously imposed mechanisms to check for endogeneity effects. Impressively, the observed subject behaviors in both sets of experiments tell similar stories. In particular, in both of them (a) costly formal sanctions cause subjects to prefer to use informal ones, (b) the latter prove to be quite efficiency-enhancing when voted for, (c) being chosen by vote appears to enhance the efficiency of using informal sanctions,⁴² and (d) there is evidence of the presence of perverse punishment and that being a perverse punisher or a recipient of perverse punishment affects voting for institutions (e.g., receiving perverse punishment increases the likelihood one votes for formal over informal sanctions). The experiments also differ in important ways, so each can also be viewed as providing an important robustness test with respect to the other, with the congruent findings of each lending support to the conclusions of the other.

One design difference is that while treatments in our paper use fixed costs of 0 or 5 points per period under the formal sanctions regime, the treatments in Markussen *et al.* have fixed costs of either 2 or 8 points per period. By considering the two sets of results side by side with one another, we observe subject choice of mechanism when using formal sanctions costs 0%, 10%, 25%, and 40% of the earnings gain associated with a shift from zero to full cooperation. As in the present paper, formal sanctions are popular in Markussen *et al.* when their cost is low, but their result tells us that that property extends to a nonzero cost (2 points or 10% of the gain from full cooperation). And as in the present paper, formal sanctions are unpopular (popular) in Markussen *et al.* when their cost is high (low), with their study adding a data point at the higher cost of 8 (2).⁴³

⁴² In fact, Markussen *et al.* obtain a more statistically significant result on this point.

⁴³ The cost comparison just given is slightly oversimplified because when formal sanctions are actually given, there is an additional cost to the sanctioned individual, and in the present paper but not Markussen *et al.* there is a corresponding cost to each remaining group member, set at 1/15 of the cost to the sanctioned individual. In practice the fixed cost constitutes by far the larger share of the cost of FS, especially when sanctions are set at a deterrent level and thus violations are rare.

Another important difference between the designs studied in the two papers is that in our paper, play under a formal sanction scheme follows both a vote for formal over informal sanctions and two votes to determine (a) which allocation is sanctioned and (b) at what rate. Under some conditions, such a sequence of votes might induce strategic voting in the first and/or second steps of the voting process, and voting between informal and formal sanctions might well be affected by uncertainty over exactly what kind of formal sanctions the group will proceed to select. Markussen *et al.* adopt a simpler design in which all details of the formal sanction regime are specified by the experimenter, so there is no chance of the votes between schemes being affected by either strategic thinking or uncertainty regarding the formal scheme's ultimate features. The fact that Markussen *et al.* obtain qualitatively similar results in the absence of those concerns suggests that our results are not dramatically affected by such problems. On the other hand, our finding that groups generally choose efficient penalties when afforded the opportunity to do so reconfirms the finding of Putterman *et al.* (2010) that subjects have the capacity to design effective schemes on their own, but does so in an environment where using an informal scheme in the first place is determined by the subjects, thus providing a result original to our paper.

A third difference between the two studies is that $\frac{1}{3}$ subjects in Markussen *et al.* engage in voting not only between formal and informal sanction mechanisms but also between each mechanism and the option of engaging in the voluntary contribution interaction without any sanctioning mechanism (equivalent to our NS condition). Their qualitatively similar outcomes regarding the choice of formal vs. informal schemes thus demonstrate robustness to presence of an additional option, while they provide novel results on choice when no sanctions play is an option. However, their design allows only two match-ups of formal vs. informal sanctions for each subject group, so our result complements theirs with more information about durability of subjects' voting choices on the issue of formal vs. informal sanctions.⁴⁴ Comparing our results to their design shows,

⁴⁴ Each subject group in Markussen *et al.* plays six four-period phases under voted regimes following a short phase of play with no scheme. In each treatment, the first vote is between informal and no sanctions, the second between formal and no sanctions, the third between formal and informal sanctions, with this sequence being repeated a second time. One interesting finding is that, similar to Gülerk *et al.* (2006) and Ertan *et al.* (2009), choice of informal over no sanctions increases in popularity with experience.

further, that presence or absence of a conditional contribution task appears to make little or no difference to subsequent play.⁴⁵

As mentioned in the introduction, it is useful to view our results also from a broader perspective that recognizes the artificiality or at least contingency of having well-functioning formal sanctioning schemes as well as perfectly enforced voting mechanisms available. That is, democracy and state machinery appear in our experimental design as a kind of *deus ex machina* or *manna from heaven* that somehow circumvent problems of collective action. But it might be best to view real world states, especially ones that are democratically accountable to their citizens (and likewise for organizational leaderships accountable to members, etc.), as manifestations of collective action problems that have already been solved, or of collective action problems needing to be addressed in ongoing ways by voluntary participation, monitoring, and other forms of engagement by citizens. While an effective state or organizational structure can be taken as exogenously given from the short-run standpoint, it is arguably an impossibility if voluntary collective action is not possible. Evidence that voluntary collective action actually does take place, then, including voluntary collective action aided by informal sanctions and by the non-trivial presence of conditional inclinations to cooperate and to punish non-cooperators, is relevant to the very possibility of formal sanctions.

There are nevertheless real world situations in which the choice between formal or informal sanctions in the more literal sense of our design has relevance. One example is the imposition of sanctions for overuse of common resources like forests and lakes, where

⁴⁵ That is, subjects in Markussen *et al.* are not exposed to a conditional contribution task but their behaviors show no qualitative difference, allaying concerns that the task may have a framing or priming effect. Finally, their design permits a more extensive and systematic investigation of play under formal sanctions that are non-deterrent in strength, an important topic given that the backing of laws by non-deterrent sanctions is probably more the rule than the exception in civilized societies. In their design, the sanction level is set exogenously, and two of their four treatments impose a sanction rate of 0.4, which ought not to influence contribution decisions if subjects are strictly self-interested and know that of one another. As can already be seen from the limited number of cases in the present paper in which a group selected a sanction rate of 0.4, the prediction that non-deterrent sanctions will be ineffective is not born out in practice. However, groups in our own experiment tend to leave the 0.4 rate for higher rates when offered the chance, whereas Markussen *et al.*'s subjects, not having that opportunity, can only vote for the informal alternative or for an NS regime when one of those is on offer. While doing so is often subjects' choice, their experiment nonetheless generates more observations of play under non-deterrent sanctions than does ours. By conducting also an exogenous non-deterrent sanction treatment, they provide evidence (paralleling Tyran and Feld, 2006) that the effectiveness of such sanctions is greatly enhanced by being chosen by vote.

limited government manpower or ambivalent views of government might make local enforcement with buy-in by the local population the better option. Another is a private voluntary organization that may choose between funding itself through entirely voluntary contributions, of which more might be induced by way of social pressures, or instead to set membership dues and punish those engaging in non-payment by means of non-provision of services and, in the limit, expulsion from the organization. A third is work teams in which effort might be elicited, on the one hand, by mutual monitoring and peer pressure or, on the other, by monitoring and penalties imposed by centralized management. These examples show that numerous details may be critical for real world application. For instance, levels of governmental capacity and popular perceptions of government may be key in the first example, or the possibility of establishing a mixed regime of formal and informal sanctions might be relevant in the second. Our experiment should therefore be seen as offering only a first, highly stylized pass at the real world problems that are at issue.

Directions for future research are numerous. Our experimental treatments modeled formal sanctions as being imposed with certainty when triggered by specified actions, but most real-world sanction schemes are fallible, both with respect to catching and imposing penalties on violators and also the possibility of mistakenly punishing non-violators. A future experiment could study subjects' willingness to pay to improve monitoring and reduce errors, as well as the effect of uncertainty of penalization and errors on choice of scheme. Another relevant issue is that it may be very difficult in practice to prevent individuals from engaging in informal sanctioning activities, in which case not only might the option of a no sanctions regime studied by Markussen *et al.* be merely hypothetical, but also the most relevant choices might be (a) informal sanctions only, and (b) formal sanctions accompanied by whatever informal sanctions individuals also decide to engage in. Recently, Kube and Traxler (2010) have studied the interaction of informal with formal sanctions in experiments in which their availability is a treatment variable. How subjects would vote on the availability of formal sanctions if informal sanctions remain available when the formal scheme is chosen might be studied in future experiments. Finally, experiments might be designed wherein the availability of a formal sanction mechanism depends on some prior voluntary collective action, a way of investigating the creation and

sustaining of state capacity to solve collective action problems when the state is understood to be the solution of a collective action problem in its own right.

References

- Andreoni, James, 1988, "Why Free Ride? Strategies and Learning in Public Goods Experiments," *Journal of Public Economics* 37: 291-304.
- Botelho, Anabela, Glenn Harrison, Ligia M. Costa Pinto and Elisabet E. Rutström, 2005, "Social Norms and Social Choice," unpublished paper, Dept. of Economics, University of Central Florida.
- Buchanan, James and Gordon Tullock, 1962, *The Calculus of Consent. Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.
- Chaudhuri, Ananish, forthcoming, "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature," *Experimental Economics* (in press).
- Cinyabuguma, Matthias, Talbot Page and Louis Putterman, 2004, "On Perverse and Second-Order Punishment in Public Goods Experiments with Decentralized Sanctioning," Brown University Department of Economics Working Paper 2004-12.
- Cinyabuguma, Matthias, Talbot Page and Louis Putterman, 2006, "Can Second-Order Punishment Deter Perverse Punishment?" *Experimental Economics* 9: 265-79, 2006.
- Dal Bó, Pedro, Andrew Foster and Louis Putterman, 2010, "Institutions and Behavior: Experimental Evidence on the Effects of Democracy," *American Economic Review* 100 (5): 2205 – 29.
- Denant-Boemont, Laurent, David Masclet and Charles Noussair, 2007, "Punishment, Counter-punishment and Sanction Enforcement in a Social Dilemma Experiment," *Economic Theory* 33: 145-167.
- Ertan, Arhan, Talbot Page and Louis Putterman, 2009, "Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free-Rider Problem" *European Economic Review* 53: 495-511, 2009.
- Falk, Armin, Ernst Fehr and Urs Fischbacher, 2005, "Driving Forces Behind Informal Sanctions," *Econometrica* 73(6): 2017-30.
- Fehr, Ernst and Simon Gächter, 2000, "Cooperation and Punishment," *American Economic Review* 90 (4): 980-94.
- Fischbacher, Urs, 2007, "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics* 10(2): 171-78.
- Fischbacher, Urs, Simon Gächter, and Ernst Fehr, 2001, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment," *Economics Letters* 71(3): 397-404.

Fischbacher, Urs and Simon Gächter, 2010, "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments," *American Economic Review* 100(1): 541-56.

Gächter, Simon and Benedikt Herrmann, 2008, "Reciprocity, Culture and Human Cooperation: Previous Insights and a New Cross-Cultural Experiment," *Philosophical Transactions of the Royal Society B* doi:10.1098/rstb.2008.0275.

Gächter, Simon, Elke Renner and Martin Sefton, 2008, "The Long-Run Benefits of Punishment," *Science* 322(Dec. 5): 1510.

Gintis, Herbert, 2000, *Game Theory Evolving*. Princeton: Princeton University Press.

Gürerk, Özgür, B. Irlenbusch and B. Rockenbach, 2006, "The Competitive Advantage of Sanctioning Institutions," *Science* 312: 108-110, April 7 2006.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter, 2008, "Antisocial Punishment across Societies," *Science* 319: 1362-1367.

Hobbes, Thomas, 1996 [1651], *Leviathan. Or the Matter, Forme and Power of a Commonwealth Ecclesiastical and Civil*. New York: Oxford University Press.

Kube, Sebastian and Christian Traxler, 2010, "The Interaction of Legal and Social Norm Enforcement," CESifo Working Paper No. 3091.

Locke, John, 2005 [1739], *Two Treatises of Government and a Letter Concerning Toleration*. Digireads.com Publishing, Stilwell.

Markussen, Thomas, Louis Putterman and Jean-Robert Tyran, 2011, "Self-Organization for Collective Action: An Experimental Study of Voting on Formal, Informal, and No Sanction Regimes," Department of Economics Working Paper 2011-4, Brown University.

Masclet, David, Charles Noussair, S Tucker and Marie-Claire Villeval, 2003, "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism," *American Economic Review* 93(1): 366-80.

Nikiforakis, Nikos, 2008, "Punishment and Counter-punishment in Public Goods Games: Can we Really Govern Ourselves?" *Journal of Public Economics* 92: 91-112.

Nikiforakis, Nikos and Hans-Theo Normann, 2008, "A Comparative Statics Analysis of Punishment in Public Goods Experiments," *Experimental Economics* 11: 358-369.

Önes, Umut and Louis Putterman, 2007, "The Ecology of Collective Action: A Public Goods and Sanctions Experiment with Controlled Group Formation," *Journal of Economic Behavior and Organization* 62: 495-521.

Ostrom, Elinor, James Walker and Roy Gardner. 1992, "Covenants with and without a Sword: Self Governance is Possible." *American Political Science Review* 86(2): 404-416.

Page, Talbot, Louis Putterman and Bulent Unel, 2005, "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency," *Economic Journal* 115: 1032-53.

Putterman, Louis, Jean-Robert Tyran and Kenju Kamei, 2010, "Public Goods and Voting on Formal Sanction Schemes: An Experiment," Working Paper, Brown University and University of Copenhagen.

Sefton, Martin, Robert Shupp and James Walker, 2007, "The Effect of Rewards and Sanctions in Provision of Public Goods," *Economic Inquiry* 45: 671-690.

Selten, Reinhard, 1975, "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory* 4: 25 – 55.

Séréville, Chantal and Bernard Myers, 1994, *Nøglen til Success I alle IQ-tester [The Key to Success in all IQ Tests (Danish)]*. Viborg: Nørhavn A/S.

Sutter, Matthias, Stefan Haigner and Martin Kocher, 2010, "Choosing the Carrot or the Stick? – Endogenous Institutional Choice in Social Dilemma Situations," *Review of Economic Studies* 77(4): 1540-66.

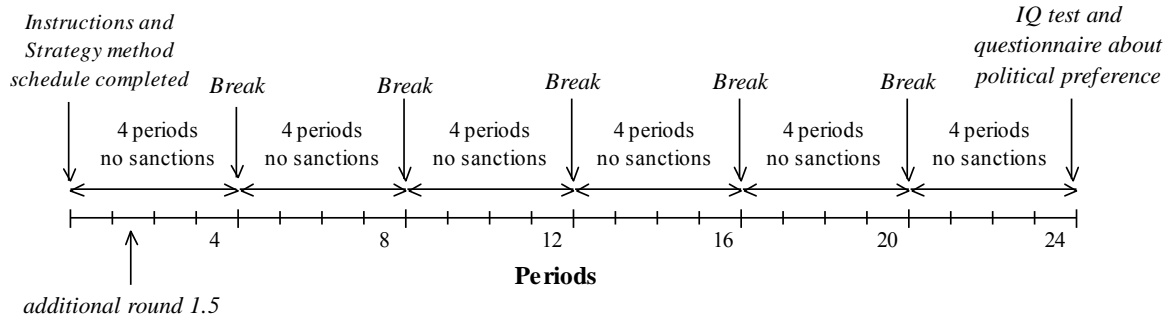
Thöni, Christian, Jean-Robert Tyran and Erik Wengström, 2009, "Microfoundations of Social Capital," working paper 09-24, Department of Economics, University of Copenhagen, October.

Tyran, Jean-Robert and Lars P. Feld, 2006, "Achieving Compliance when Legal Sanctions are Non-deterrent," *Scandinavian Journal of Economics* 108(1): 135-56.

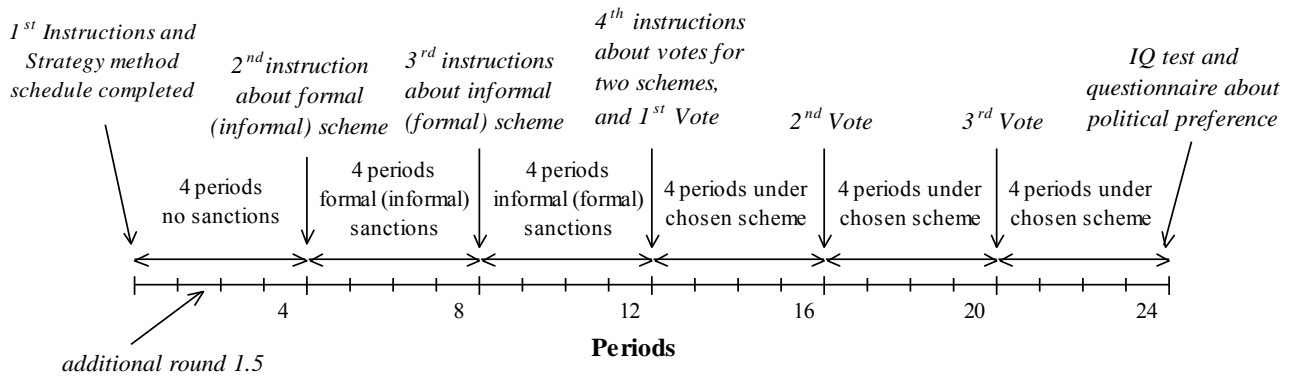
Zelmer, Jennifer, 2003, "Linear Public Goods Experiments: A Meta-Analysis," *Experimental Economics* 6: 299-310.

Figure 1: Experimental Design

(A) BASELINE treatment



(B) 3 vote treatments



(C) 6 vote treatments

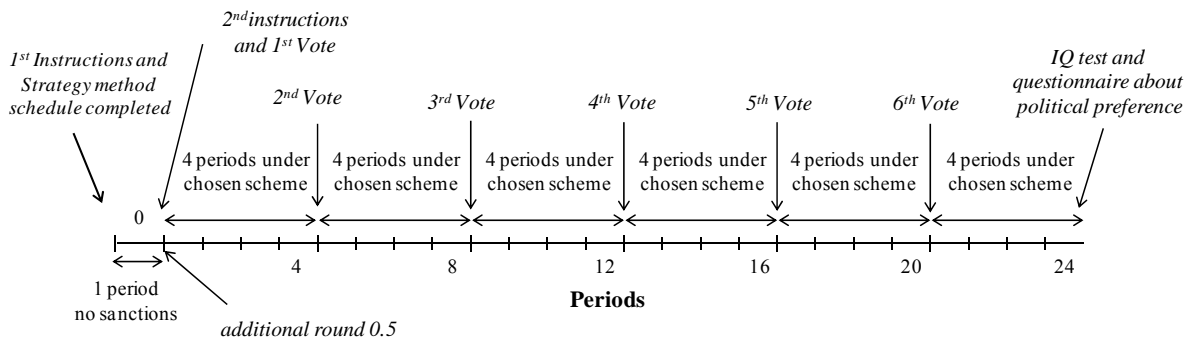
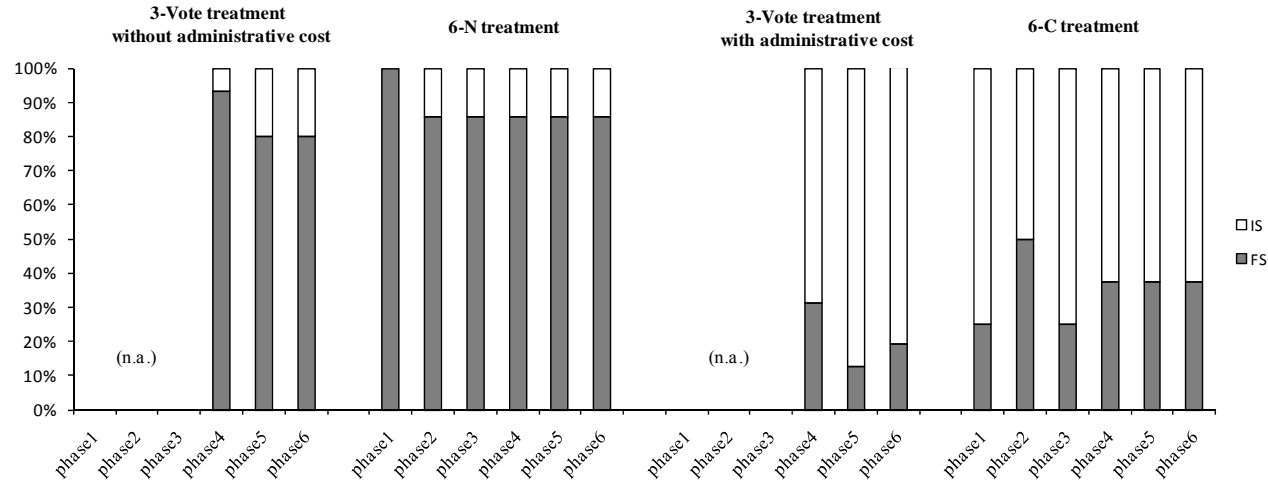
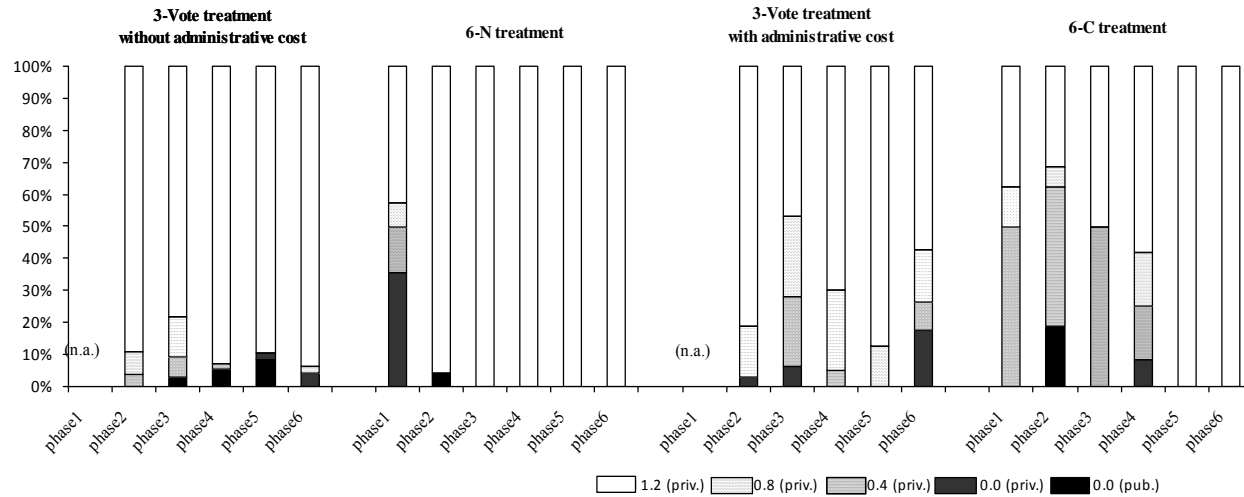


Figure 2: The share of formal vs. informal schemes, and the shares of formal scheme at different sanction rates, by phase



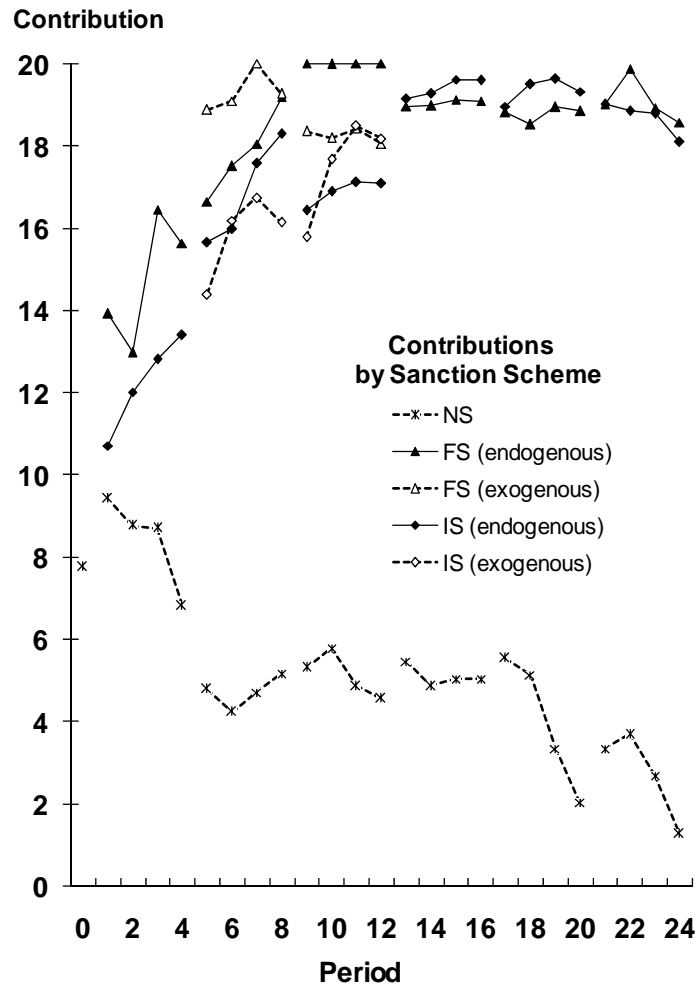
(a) The shares of formal scheme vs. informal scheme



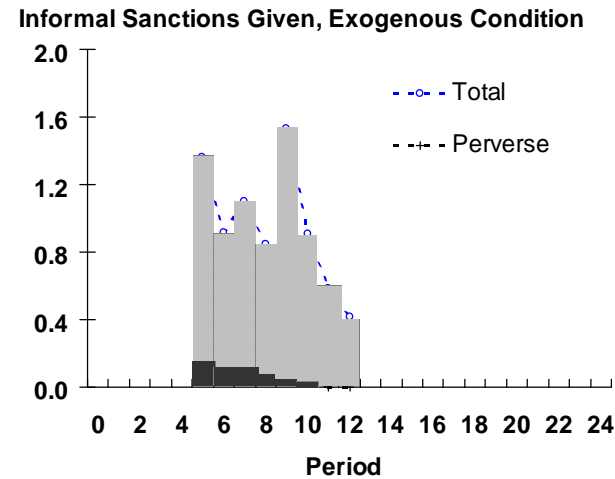
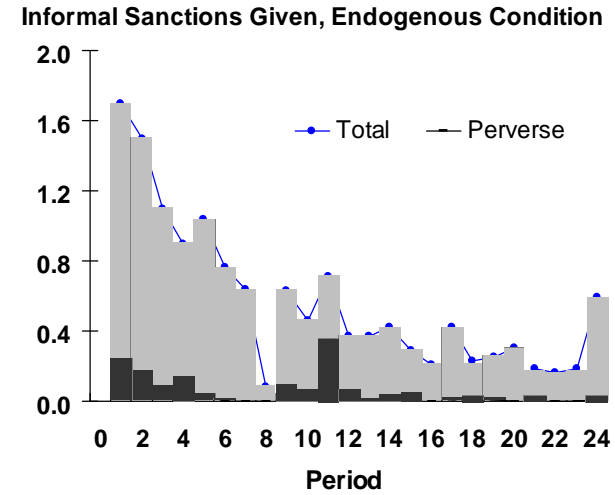
(b) The shares of FS at different penalty rates

Note: data are group outcomes, not individual votes.

Figure 3: The trends of average contribution to the public account and average amount of informal sanctions given

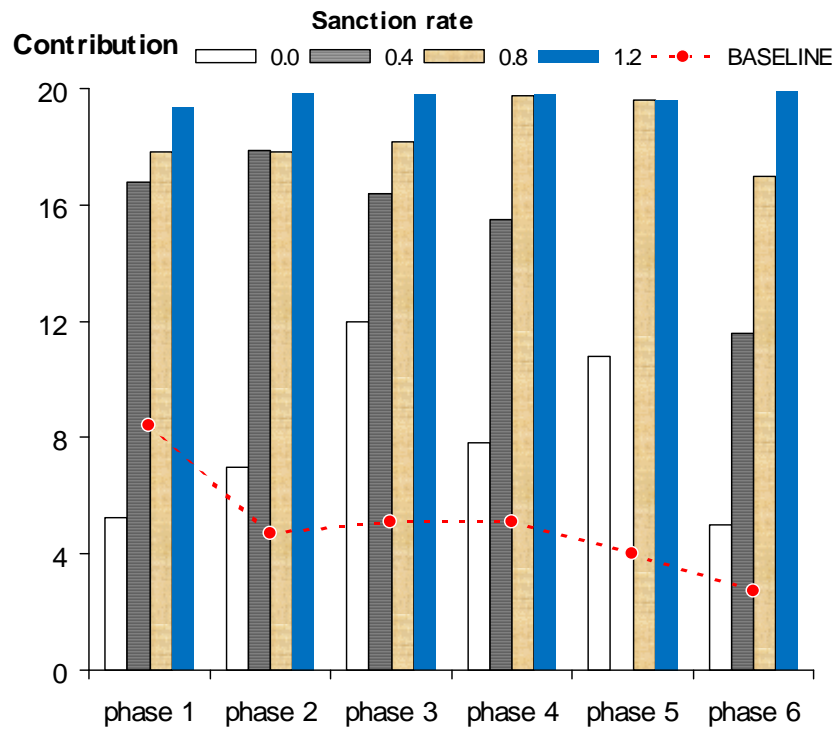


(a) Average Contribution



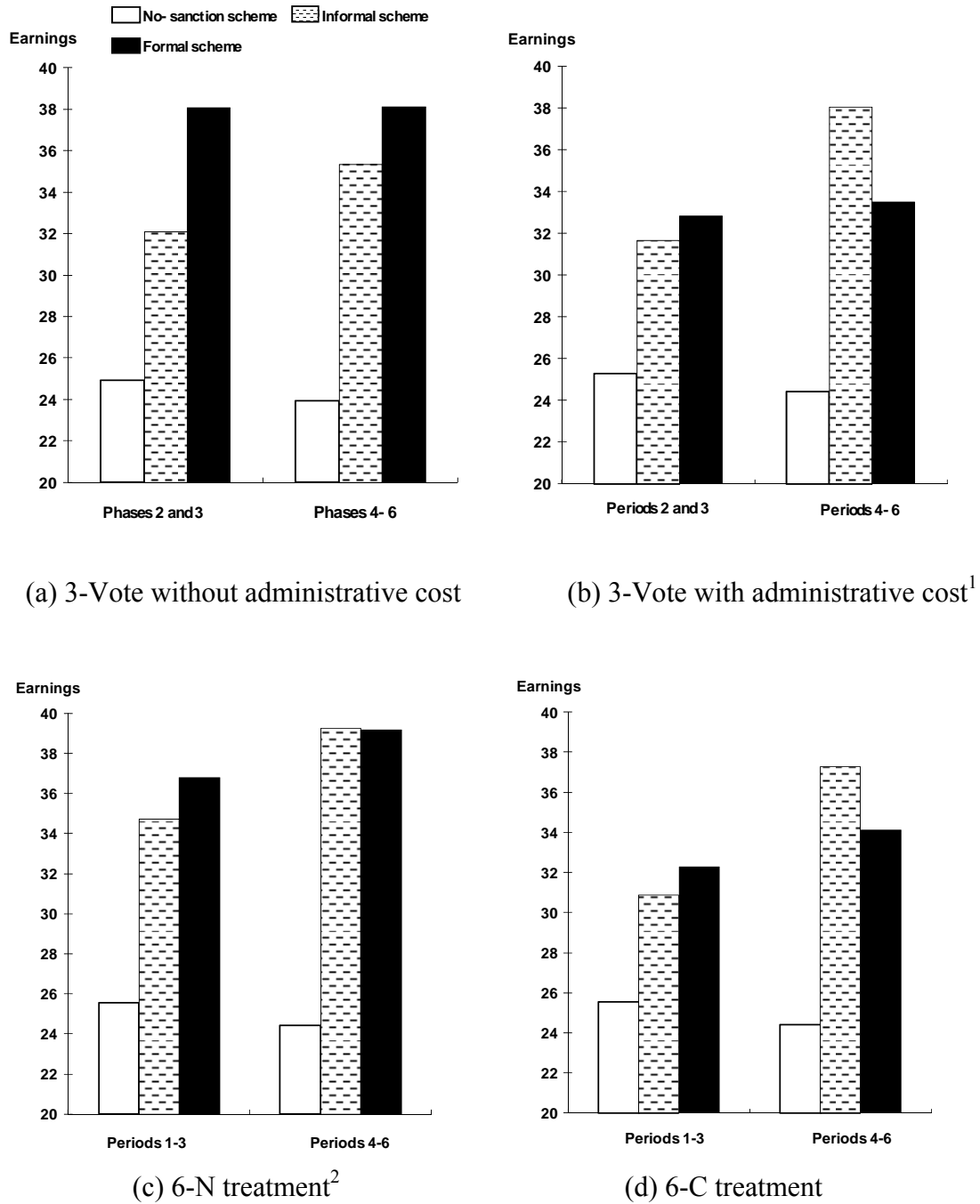
(b) Average Amounts of Informal Sanction

Figure 4: Contributions under formal scheme by sanction rates



Note: No group chose a sanction rate of 0.4 in phase 5.

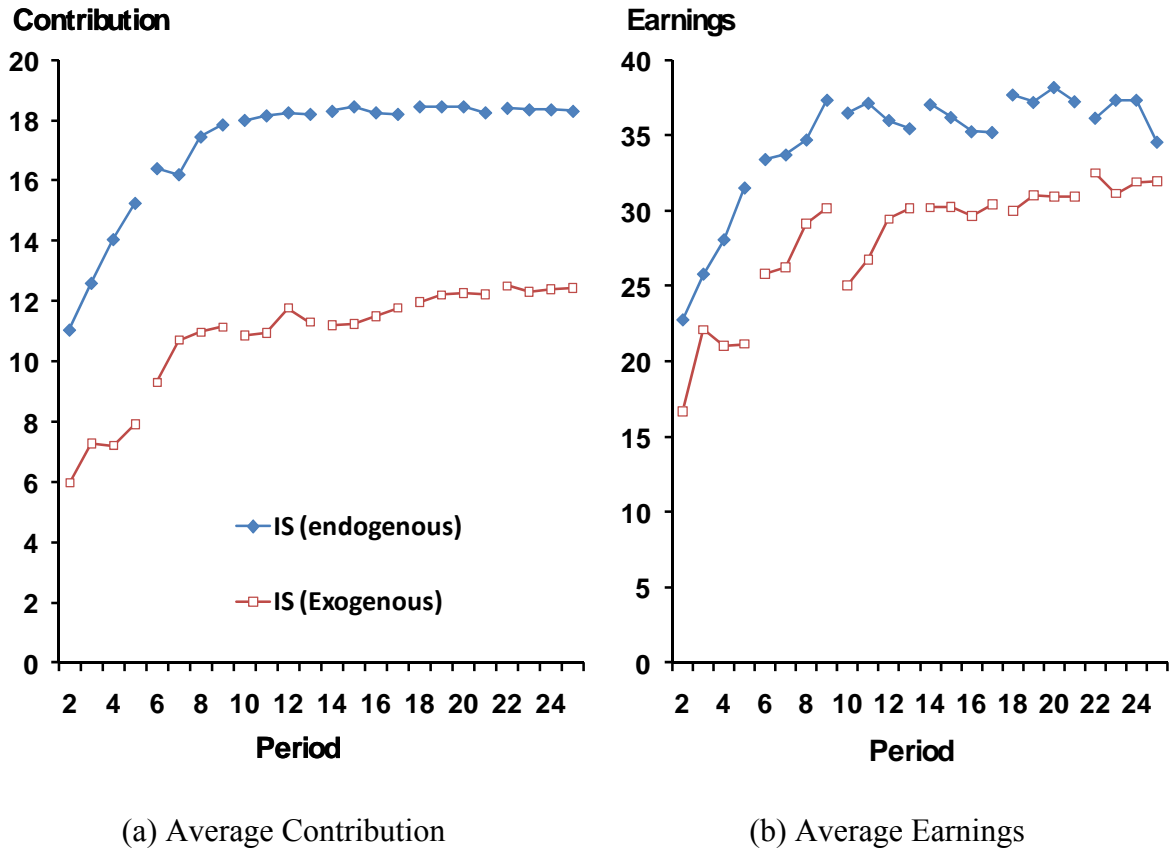
Figure 5: Average earnings under NS, IS, and FS in early and late phases



Notes:

1. In the 3-Vote with administrative cost, in phase 6 only 3 groups chose formal scheme, and one of them chose nonbinding sanction rate in periods 21 and 24 (See Appendix Table B.1).
2. In the 6-N treatment, no (1) group chose informal scheme in phase 1 (in phases 2 – 6).

Figure 6: Contributions and earnings in 6-N treatment groups selecting IS and in test treatment groups exogenously assigned IS



Note: For Phase 1, dashed line and circles indicate the average for all five groups selecting IS endogenous, while the solid line with diamonds indicates the average for the four groups that continued to select IS in all phases.

Table 1: Summary of treatments, sessions, and subjects

		No Fixed Cost of FS (N)	Fixed Cost of FS (C)
3-Vote Treatments	NS→IS→FS	Treatment: 3-(IF)-N Sessions: 2 Groups: 8 Subjects: 40	Treatment: 3-(IF)-C Sessions: 2 Groups: 8 Subjects: 40
	NS→FS→IS	Treatment: 3-(FI)-N Sessions: 2 Groups: 7 Subjects: 35	Treatment: 3-(FI)-C Sessions: 2 Groups: 8 Subjects: 40
6-Vote Treatments		Treatment: 6-N Sessions: 2 Groups: 7 Subjects: 35	Treatment: 6-C Sessions: 2 Groups: 8 Subjects: 40 <div style="border: 1px dashed black; padding: 5px; margin-top: 10px;"> Exogenous IS Comparison Treatment Sessions: 2 Groups: 6 Subjects: 30 </div>
Baseline (No Sanctions)		Treatment: Baseline (Exogenous-NS) Sessions: 2 Groups: 8 Subjects: 40	

Notes: NS = no sanctions, FS = formal sanctions, IS = informal sanctions. The experiment as a whole consisted of 16 sessions in which 60 groups composed of 300 individual subjects participated. There were a total of 183 group votes on the use of FS versus IS, with 915 individual votes on that question vote. The Exogenous IS comparison treatment (dashed rectangle) appears in the cell of Treatment **6-C** because its parameters and structure are identical to that of groups selecting IS in **6-C** treatment except for the fact that the IS scheme is imposed rather than chosen by vote.

Table 2: Percentage of group outcomes and individual votes for formal (vs. informal) sanction scheme

Treatment	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phases 4-6
3(FI)-N	n.a.	n.a.	n.a.	86, 71	71, 63	71, 63	76, 66
3(IF)-N	n.a.	n.a.	n.a.	100, 85	88, 75	88, 78	92, 79
6-N	100, 80	86, 80	86, 80	86, 71	86, 74	86, 74	86, 73
Overall, no admin cost	100, 80	86, 77	86, 80	91, 76	82, 71	82, 72	85, 73
3(FI)-C	n.a.	n.a.	n.a.	25, 33	13, 23	25, 20	21, 25
3(IF)-C	n.a.	n.a.	n.a.	38, 48	13, 23	13, 33	21, 36
6-C	25, 43	50, 48	25, 33	38, 35	38, 38	38, 30	38, 34
Overall, admin Cost	25, 43	50, 48	25, 33	33, 33	17, 27	25, 26	25, 29
Overall	60, 60	67, 61	53, 55	61, 54	50, 48	52, 48	54, 50

Notes: The pairs of numbers in each cell of this table indicate the percentage of groups choosing FS (left) and the percentage of individuals voting for FS (right). The remaining votes (i.e. 100% - number indicated) are for IS. See Table 1 for the absolute number of groups and individuals in the respective treatment.

Table 3: Determinants of Votes between Formal and Informal Schemes in 3-Vote Treatments

(a) 3(FI)-N and 3(IF)-N, pooled

Independent variable	(1)	(2)	(3)	(4)	(5)
(Earning under formal scheme) /(earnings under informal scheme)	2.94** (1.19)	3.08** (1.27)	5.29*** (1.71)	2.18** (1.16)	3.96** (1.57)
(CV ^a under formal scheme)/(CV under informal scheme)	---	0.038 (0.10)	0.069 (0.10)	0.020 (0.10)	0.046 (0.10)
Gave Perverse Punishment	---	---	-2.06** (0.92)	---	-1.83** (0.84)
Received Perverse Punishment	---	---	-2.28** (1.07)	---	-1.55 (0.97)
Avg. Conditional Contribution	---	---	---	0.0013 (0.10)	0.016 (0.040)
IQ	---	---	---	0.28*** (0.10)	0.22** (0.096)
Gender	---	---	---	0.12 (0.44)	0.15 (0.43)
General Political Orientation	---	---	---	0.22 (0.16)	0.25 (0.16)
Vote number	-0.22 (0.15)	-0.23 (0.16)	-0.23 (0.16)	-0.23 (0.16)	-0.23 (0.16)
Phase 2 dummy	-0.21 (0.48)	-0.20 (0.48)	-0.65 (0.50)	-0.23 (0.44)	-0.56 (0.45)
Constant	-1.86 (1.50)	-2.03 (1.59)	-4.07** (1.91)	-3.66** (1.72)	-5.15** (1.98)
Number of observations	225	225	225	225	225
Log likelihood	-103.8	-103.7	-98.5	-97.5	-93.7
Wald chi ²	9.13	9.01	13.9	16.3	18.7
Prob > chi ²	.028	.061	.031	.039	.043

(b) 3(FI)-C and 3(IF)-C, pooled

Independent variable	(1)	(2)	(3)	(4)	(5)
(Earning under formal scheme)/(earnings under informal scheme)	1.61*** (0.46)	1.41*** (0.48)	1.16*** (0.52)	1.44*** (0.46)	1.11** (0.51)
(CV under formal scheme)/(CV under informal scheme)	---	-0.15 (0.32)	-0.16 (0.33)	-0.36 (0.33)	-0.36 (0.34)
Perv Pun Dummy	---	---	0.72 (0.81)	---	0.78 (0.77)
Perv Rec Dummy	---	---	0.61 (0.62)	---	0.79 (0.61)
Conditional Contribution	---	---	---	0.0056 (0.040)	-0.011 (0.043)
IQ	---	---	---	-0.18** (0.072)	-0.19** (0.074)
Gender	---	---	---	0.47 (0.36)	0.58 (0.38)
General Political Orientation	---	---	---	-0.091 (0.12)	-0.077 (0.12)
Vote number	-0.16 (0.14)	-0.16 (0.14)	-0.18 (0.14)	-0.16 (0.14)	-0.18 (0.14)
Phase 2 dummy	-0.41 (0.38)	-0.24 (0.38)	-0.12 (0.40)	-0.33 (0.36)	-0.19 (0.38)
Constant	-2.26*** (0.63)	-1.95*** (0.72)	-1.82*** (0.73)	-0.57 (0.96)	-0.35 (0.98)
Number of observations	240	225 ^b	225 ^b	225 ^b	225 ^b
Log likelihood	-111.5	-109.4	-108.7	-104.2	-103.0
Wald chi ²	14.6	12.8	13.3	20.4	20.2
Prob > chi ²	.002	.012	.039	.009	.028

Notes: 1. Random-effects Probit regressions. The dependent variable equals 1 if the subject voted for the use of the formal regime, 0 otherwise.

2. The Perv Pun Dummy equals 1 if a subject has ever punished another member even if the member contributed above the median of contributions in a group under exogenous informal scheme, 0 otherwise. The Perv Rec Dummy equals 1 if a subject has received punishment even if the subject contributed above the median of contributions in a group under exogenous informal scheme, 0 otherwise.

3. The Female variable equals 1 if a subject is female, 0 otherwise.

4. Phase 2 dummy equals 1 if Phase 2 is under formal scheme, 0 otherwise. Vote number $\in \{1, 2, 3\}$.

5. The figure in the parenthesis is standard deviation.

6. *, **, and *** indicate significance at the .10 level, at the 0.05 level and at the .01 level, respectively.

a. CV stands for the coefficient of variation.

b. 15 observations in which CV under informal scheme equals 0 are dropped.